DOCUMENT RESUME

ED 294 738 SE 049 132

AUTHOR	Shavelson, Richard; And Others
TITLE	Indicator Systems for Monitoring Mathematics and Science Education.
INSTITUTION	Rand Corp., Santa Monica, Calif.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	ISBN-0-8330-0804-8; RAND-R-3570-FSF
PUB DATE	Aug 87
GRANT	SPA-8470440
NOTE	96p.; Charts may not reproduce well.
AVAILABLE FROM	Publications Department, The RAND Corporation, 1700 Main Street, Santa Monica, CA 90406-2138 ($7.50, 25 or more 25% discount).
PUB TYPE	Reports - Descriptive (141) -- Reports - Research/Technical (143)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Educational Change; *Educational Improvement; Educational Quality; *Elementary School Mathematics; *Elementary School Science; Elementary Secondary Education; Mathematics Education; Science Education; *Secondary School Mathematics; *Secondary School Science
IDENTIFIERS	Indicators; *National Science Foundation; *Science Education Research

ABSTRACT

The demand for more valid and useful educational indicators has grown significantly as national, state, and local agencies have moved to improve the quality of elementary and secondary education. At the national level, there is a growing need for more systematic information about student performance in mathematics and science, about the content of courses taught, and about the quality of mathematics and science teachers. At the state and local levels, increased academic standards have led to a need for more sophisticated measures of the processes and outcomes of schooling. No comprehensive indicator system is available to measure the status of mathematics and science education in the United States. This report addresses the question of the National Science Foundation's role in monitoring mathematics and science education. The report identifies several options for developing a system of education indicators, i.e., measures that report the condition of particularly significant features of mathematics and science education, and then assesses each option in terms of its usefulness, feasibility, and compatibility with other national and state efforts to monitor educational performance. (CW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294738

Indicator Systems for Monitoring Mathematics and Science Education

Richard Shavelson, Lorraine McDonnell,
Jeannie Oakes, Neil Carey
With Larry Picus

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Moira Allen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SE 049 132

RAND

BEST COPY AVAILABLE

The work described in this report was supported by the National Science Foundation under Grant No. SPA-8470440.

ISBN: 0-8330-0804-8

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of The RAND Corporation do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

R-3570-NSF

Indicator Systems for Monitoring Mathematics and Science Education

Richard Shavelson, Lorraine McDonnell,
Jeannie Oakes, Neil Carey
With Larry Picus

August 1987

Supported by the
National Science Foundation

RAND

PREFACE

Growing concern about the level of scientific literacy and the future availability of scientific manpower has led to a need for more systematic information about participation and achievement in mathematics and science courses, the content of such courses, and the quality of those teaching them. The National Science Board's Commission on Precollege Education in Mathematics, Science, and Technology issued a report in 1983 that set as the nation's goal "the highest quality education and highest participation level in the world by the year 1995." The report, entitled *Educating Americans for the 21st Century*, went on to recommend that the federal government "finance and maintain a national mechanism to measure student achievement and participation in a manner that allows national, state and local evaluation and comparison of education progress."

Because the National Science Foundation (NSF) has specific responsibility to strengthen mathematics and science education, the National Science Board recommended that the NSF monitor the progress of education in these subject areas. Yet despite this clear charge, it was not obvious what the NSF should do, given that other agencies already collect and report education statistics.

The RAND Corporation undertook a study to identify a set of indicator systems for monitoring precollege mathematics and science education; to analyze the benefits and problems associated with each option; and to ascertain whether, on balance, one or more indicator systems could be justified and if so, what they might cost. The work was supported under NSF Grant SPA-8470440, 86-GA-0153. This report describes the study background, conceptual framework, methods, and findings.

This report is intended primarily to assist the NSF in developing its role in monitoring precollege science education. However, it should also be of interest to other federal agencies and policymakers charged with monitoring the nation's educational progress, as well as to state and local policymakers and educators. Finally, it should be useful to the many mathematics and science education constituencies responsible for improving the quality of teaching and the achievements of students in this vital area.

SUMMARY

The demand for more valid and useful educational indicators has grown significantly as national, state, and local agencies have moved to improve the quality of elementary and secondary education. At the national level, there is a growing need for more systematic information about student performance in mathematics and science, about the content of courses taught, and about the quality of mathematics and science teachers. At the state and local levels, increased academic standards have led to a need for much more sophisticated measures of the processes and outcomes of schooling. Although data on student and teacher characteristics, the schooling process, and student achievement are now collected, no comprehensive indicator system is available to measure the status of mathematics and science education in the United States.

Data from national assessments of science and mathematics and other sources such as Scholastic Aptitude Test (SAT) scores have signaled that science and mathematics education face serious difficulties in the United States. However, the available data do not allow policymakers to identify the sources of these problems with sufficient precision to address them effectively. In 1983, the National Science Board (NSB) recommended that the National Science Foundation (NSF) monitor the progress of science and mathematics education.

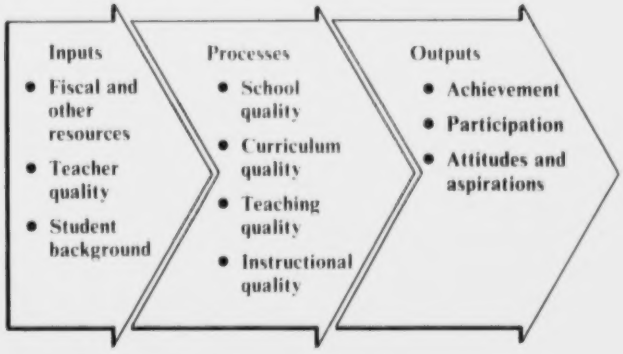
This report addresses the question of the NSF's role in monitoring mathematics and science education. The report identifies several options for developing a system of education indicators, i.e., measures that report the condition of particularly significant features of mathematics and science education, and then assesses each option in terms of its usefulness, feasibility, and compatibility with other national and state efforts to monitor educational performance.

We recommend that the NSF should initially construct a patchwork of existing indicators to describe the status of mathematics and science education. At the same time, it should undertake developmental research to create better indicators than currently exist, work toward an institutionalized method of collaborating with other data collection efforts, and establish procedures for ensuring that monitoring system results can be used constructively by policymakers and educators.

THE USE OF AN INDICATOR SYSTEM TO MONITOR EDUCATION

Fundamental to our conception of monitoring is the notion of indicators. Education indicators should tell a great deal about the entire educational system by reporting on individual features of that system. Such indicators are the basic unit for monitoring the progress of schools.

To monitor a system accurately, indicators must be derived from a sound conceptual model of how that system actually works. Current research about schooling is not sufficiently advanced to support a strictly predictive or causal model, but it can provide a framework that includes the central features of mathematics and science education and identifies the logical relationships among those features. An extensive review of the literature on social indicators and educational research provided a basis for formulating a model of the education system and the indicators for measuring each component. This model contains inputs (the human and financial resources available to the education system), processes (what is taught and how it is taught), and outputs (the consequences of schooling on students from different backgrounds):



Our research review also suggested how various elements of the system are likely to be logically or empirically related. Based on that review, interviews with policymakers, and ongoing consultation with the national advisory board for the project, we identified the following criteria for selecting elements of the schooling process that should be incorporated in an indicator system. A valid indicator should:

- Provide information that describes central features of the educational system.
- Provide information about current or potential problems.
- Provide policy-relevant information.
- Measure observed behavior rather than perceptions.
- Provide analytical links among important components of schooling.
- Generate data from measures that are generally accepted as valid and reliable.
- Provide information that can be readily understood by a broad audience.
- Be feasible in terms of timeliness, cost, and expertise.

These criteria were applied to 104 potential indicators identified from past research as key measures of schooling. Forty of these indicators satisfied our criteria and thus became the core of five indicator system options.

INDICATOR SYSTEM OPTIONS

The five options are:

- *Status quo.* Under this option, the NSF would rely on whatever data were available at the time a report is produced or a policy issue considered. Status quo indicators do not represent an indicator system, since they neither stem from an overarching conception of the set of indicators needed, nor are they necessarily part of a systematic data collection effort.
- *Patchwork.* Under the patchwork option, indicators would be pieced together from existing data sources, such as the National Assessment of Educational Progress (NAEP), guided by an integrated and conceptually sound model of schooling.
- *Cyclical studies.* The NSF could commission data collection on a regular basis to provide a time series on several domains of schooling, such as achievement, teacher quality, and curriculum quality. In their emphasis on depth of understanding, cyclical studies would typically complement whichever indicator options the NSF might select. For that reason, they would be based on smaller samples and could explore topics in ways that a large-scale data collection effort could not.
- *Piggyback.* The NSF could buy into ongoing data collection programs to obtain more complete data about mathematics and science education and about areas of particular priority.

- *Independent.* Under the independent option, the NSF would develop and field a comprehensive data collection system of its own that spans the major components of schooling.

ASSESSING THE OPTIONS

Like most policy alternatives, none of our indicator system options is clearly superior to any other, so selection among them must be based on factors such as the NSF's preferences about how information should be used and the funding available for the system.

An indicator system can describe national trends (e.g., in achievement, teacher quality, and curriculum quality); describe state-by-state trends; function as an early warning system by identifying emerging problems; inform the improvement of policy and practice; and enable the NSF to provide leadership by monitoring curricular and achievement areas that are presently ignored. Only the piggyback and independent options can fully serve all these purposes. The other options are deficient in their ability to describe state-level conditions and to provide leadership in indicator development that can provide direction for more fundamental policy change. However, those options that are less useful are generally more feasible to implement.

Feasibility is a question of the level of technical expertise required, respondent burden, likely political support, extent of interagency cooperation needed, stability over time, and cost. We found that as one moves from the status quo to a fully independent indicator system, technical expertise becomes an increasingly important issue. That is, an expansion of existing data collection programs to meet the NSF's specifications (a piggyback system) or the operation of an independent system requires state-of-the-art breakthroughs in sampling, measurement, and data analysis. Likewise, the independent option increases respondent burden because it adds yet one more data collection effort in the schools. Political support tends to increase with more comprehensive indicator systems because the enhanced monitoring capacity more closely addresses policymakers' concerns and suggests possible targets for reform. However, respondent burden and political support generate very real tradeoffs, particularly at the local level: Support depends on how policymakers perceive the usefulness of monitoring information as compared with the overall burden the indicator system imposes. Interagency cooperation is especially important in the piggyback option, whereas the distribution of authority across agencies becomes an issue with an independent system. The independent option is most likely to be stable over time, since it is under the direct

control of the NSF. Annual costs for different indicator options range from the approximately $40,000 over staff costs that is currently spent to commission special analyses of existing data to the $23 million to $34 million that would be required for an independent system.

One final factor to consider in weighing different indicator options is the environment in which an NSF indicator system would be implemented. The U.S. Department of Education (ED) is currently redesigning its entire data collection and reporting system, including the NAEP. Similarly, the Council of Chief State School Officers is working to define a set of comparable indicators that would be collected by all fifty states. This suggests that opportunities for collaboration will grow significantly over the next decade, but also that any decision made by the NSF in the next few years may have to be modified as other actors design their indicator strategies.

RECOMMENDATIONS

After assessing these tradeoffs, we recommend that the NSF:

- Develop an indicator system that both describes and relates essential elements of the mathematics and science education system.
- Provide national and state-level information on key science and mathematics indicators every two years.
- Over the next three to five years, develop better indicators of mathematics and science education using existing data—by building a patchwork system that both reports the current status of the system and identifies critical gaps in existing indicators and analytical methods.
- Work to influence the direction of other indicator activities, and then build on those data collection efforts (the piggyback option) to expand the amount and quality of the data available on mathematics and science education.
- Conduct adjunct studies that are stimulated by the findings of the indicator system, but that go significantly beyond them in analyzing the causes of observed changes and in suggesting alternative policy implications.
- Develop new measures where adequate ones are not currently available. The priorities should be the design of more valid indicators of scientific achievement (including students' ability to think critically and to apply knowledge in solving problems); better measures of teacher quality; and more comprehensive information about the science and mathematics curriculum

(e.g., its depth of coverage and scientific accuracy and the process by which students are instructed).

- Develop specific procedures for analyzing and reporting indicators that ensure that results are well-reviewed (perhaps by a body such as the National Academy of Sciences), appropriately disseminated, and assessed regularly for their policy implications.

ACKNOWLEDGMENTS

Throughout this research, we benefited from the sage advice and unfailing support of Richard Berry of the Studies and Analysis Program, Science and Engineering Education Directorate, National Science Foundation. He has become the education indicator broker in Washington, serving in the role of matchmaker among diverse projects conducted by the Center for Statistics (U.S. Department of Education), the Council of Chief State School Officers, the National Academy of Sciences, the National Assessment of Educational Progress, the Wisconsin Center for Mathematics Indicators, and this project. Our project has profited greatly from our working knowledge of these other indicator efforts. Richard Berry, about to retire, has left a lasting imprint on the educational indicator movement. We wish him well, and we will greatly miss him.

The project was guided by a national advisory board, representing mathematicians and scientists, science and mathematics educators, psychometricians, policy analysts, and policymakers. We are indebted to the following members of the committee for their hard work, apt guidance, and valuable comments: Leigh Burstein, Audrey Champagne, Lee Cronbach, Pascal Forgione, Bernard Gifford, Norman Hackerman, Harry Handler, Ina Mullis, Ingram Olkin, Jerome Pine, Senta Raizen, Tom Romberg, and Ramsay Selden.

This report profited from the helpful and insightful reviews of our colleagues Anthony Pascal of RAND and Richard Murnane of Harvard. Of course, we are responsible for any errors of commission or omission. Janet DeLand sharpened our prose, and Linda Weiss and Jeanne Shultz provided patient, invaluable assistance in the production of this report. Toby O'Brien, Linda Weiss, and Susan Bell took excellent care of us and our advisory board at its meetings.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS xi

FIGURES AND TABLES xv

Section

I. INTRODUCTION 1
- Diagnosing the Problem 1
- Why Should the NSF Monitor Precollege Science and Mathematics Education? 2
- Purpose of This Study 4
- Study Methods 5

II. A COMPREHENSIVE MODEL OF THE PRECOLLEGE EDUCATIONAL SYSTEM 8
- Indicators and Indicator Systems 8
- A Comprehensive Model of Schooling 9
- Components of Schooling and Key Indicators 12
- Conclusions: Using a Comprehensive Model for Identifying Indicators 21

III. DEVELOPING A USEFUL INDICATOR SYSTEM 22
- Policymaker Interviews 23
- Criteria for Selecting Indicators 27

IV. INDICATOR SYSTEMS: FIVE GENERIC OPTIONS 30
- The Status Quo Option 30
- The Patchwork Option 31
- The Cyclical Studies Option 34
- The Piggyback Option 35
- The Independent Option 35

V. ASSESSING THE OPTIONS 39
- Usefulness 39
- Feasibility 42

VI. RECOMMENDATIONS AND NEXT STEPS 47
- Recommendations 47
- Next Steps 51
- Opportunity, Promise, and Cautions 56

Appendix
A. Central Features of Schooling 59
B. Cost Analysis 64
C. Project National Advisory Board Members 72
D. Interview Topics 73

REFERENCES 77

FIGURES

1. Trends in overall science achievement scores: 1969–1982 . . 3
2. Trends in Scholastic Aptitude Test (SAT) scores 4
3. Process for generating NSF options 6
4. Linking elements of the educational system 10
5. An example of a patchwork indicator system 32
6. An example of a piggyback indicator system 36
7. Uses of indicator options 40
8. Feasibility of indicator options 43
9. Indicators in the context of NSF's broader mission 48

TABLES

1. Patchwork indicator system 33
2. Piggyback indicator system 37
3. Cost of indicator options 45

B.1. Estimated cost of an independent system 70
B.2. Assumptions used to derive Table B.1 71

I. INTRODUCTION

DIAGNOSING THE PROBLEM

The United States is increasingly being required to make important decisions and to compete internationally in areas directly related to science and mathematics—computer technology, biological warfare, AIDS, and nuclear energy, to name just a few. Concerns about dependence on technology and technical knowledge have pushed two urgent questions to the forefront of education policy: Can the precollege educational system ensure an adequate level of scientific literacy for all students? And will our schools produce an adequate supply of students mathematically and scientifically well-educated enough to pursue advanced training in these disciplines? Many analysts are concerned that U.S. schools may not be generating the levels of scientific understanding needed to confront critical decisions, or to produce an adequately prepared workforce.[1] A wealth of data suggest that the answers to both questions are not heartening. In recent years, public education in science and mathematics has fallen short of what is both possible and necessary. The recent Second International Mathematics Study raised considerable concern about levels of general literacy, with its finding that the mathematics achievement of U.S. eighth and twelfth graders typically lags far behind that of their peers in many other nations (McKnight et al., 1987).

Other data show that since the early 1970s, fewer students have been entering scientific fields, and fewer scientific doctoral degrees have been awarded (Bloch, 1986). At the same time, between 1976 and 1983, jobs for scientists and engineers increased at three times the rate of U.S. employment generally, and the annual employment growth in computer specialities during these years was 17 percent (NSB, 1985). National policymakers have been alarmed by figures such as these, which suggest that the demand for a highly trained scientific workforce may soon outstrip the supply. Finally, there is widespread concern that the United States is not producing enough mathematics

[1]The National Science Foundation (NSF) has specific responsibility to strengthen science education at all levels, as mandated by the National Science Foundation Act, 1950; 42 U.S.C. 1961-1975. Moreover, in 1983, the National Science Board Commission on Precollege Education in Mathematics, Science, and Technology stated that a primary goal should be to provide all the nation's youth with a level of education in mathematics, science, and technology that is of the highest quality attained anywhere in the world and that reflects the particular and peculiar needs of our nation (NSB, 1983).

and science educators to fill the projected shortfalls in qualified junior high school teachers projected for the late 1980s and early 1990s, and in qualified senior high school teachers projected for the mid-1990s (Darling-Hammond, 1984).

Recognizing this growing crisis in American education, yet not having adequate data to pinpoint the possible reasons, the National Science Board (NSB) in 1983 released a study of precollege mathematics and science education calling for better instruction in science, mathematics, and technology (NSB, 1983). The report called for upgrading the quality of science and mathematics for *all* students. It strongly linked the ability of American schools to upgrade the quality of mathematics and science to the willingness of those schools to provide better instruction for students preparing for careers in scientific fields, and also to the schools' ability to better educate the general student population. The commission cautioned that attainments of *all* students are increasingly critical to the nation's scientific and technological strength.

The NSB was not alone in its concern. Similar charges were issued in reports by the U.S. Department of Education (ED) Commission on Excellence in Education, the Education Commission of the States Task Force on Education for Economic Growth, and the Twentieth Century Fund's Task Force on Federal Elementary and Secondary Education Policy. Since these early reports, other analysts have also warned that the nation's future economic and security interests hinge on improving the quality of precollege education for all students (e.g., Matlack, 1987).

However, the NSB report went even further. It also strongly recommended that the NSF undertake an effort to collect statistics on students' participation and achievement, in order to *monitor* progress toward the goal of "the highest quality education and highest participation level in the world by 1995."

WHY SHOULD THE NSF MONITOR PRECOLLEGE SCIENCE AND MATHEMATICS EDUCATION?

To understand this recommendation that the NSF take responsibility for "monitoring" educational progress in mathematics and science, we must consider the data the NSB and the nation had before them in 1983—data that continue to be of concern. National achievement-test scores had been steadily declining for more than a decade. Figure 1 shows trends in overall science achievement test scores.[2] Scholastic

[2]Some achievement data for 9-year-olds also suggest declines. However, comparable time-series data are not available for that age group.

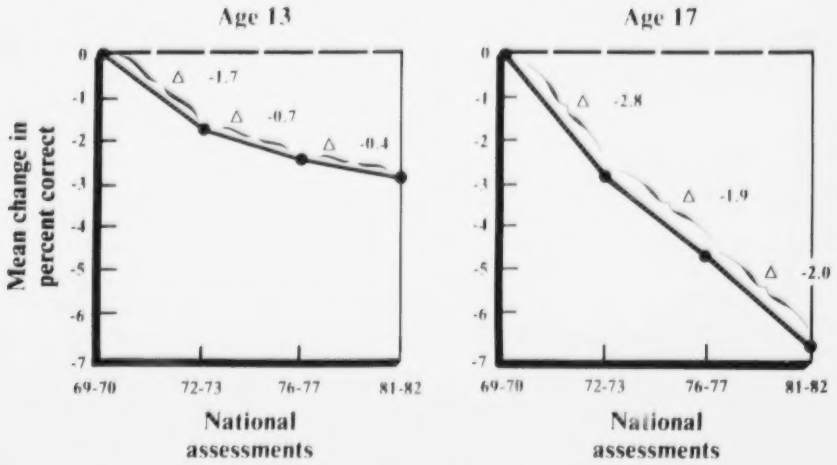


Fig. 1—Trends in overall science achievement scores: 1969–1982
(SOURCE: National Assessment of Educational Progress, Hueftle et al. (1983))

Aptitude Test (SAT) scores also have been declining since the mid-1960s, as shown in Fig. 2. These declining scores suggest that less academically able students may be entering higher education. Moreover, other statistics indicate that the top U.S. students do not perform as well as their counterparts in other countries and that U.S. educational performance levels overall have declined since the mid-1960s (McKnight et al., 1987). However, adequate data were not available to permit potential problems within the education system to be identified or to enable policymakers and educators to target reform efforts with the necessary precision.

The data we do have raise a number of questions. For example, is the decline in achievement-test scores due to a drop in the quality of the teaching force? Are high school students taking fewer mathematics and science courses than they did previously? Or is an outdated or inefficient curriculum the reason for the problem? Is the decline caused by a shift in student demographics? Are there motivational factors influencing student performance? These questions cannot be fully answered with currently available data.

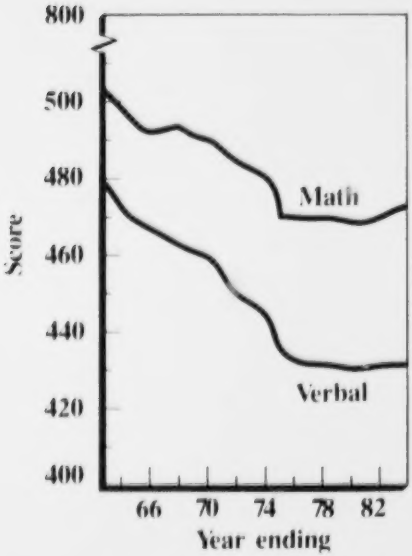


Fig. 2—Trends in Scholastic Aptitude Test (SAT) scores
(SOURCE: Admissions Testing Program of the College Board)

The NSB recognized the sorry state of education statistics and charged the NSF with monitoring the progress of science and mathematics education. The Board believed that a monitoring system could begin to provide the information necessary to chart progress toward improved precollege education and could also generate the data necessary to illuminate and interpret trends as they appeared.

PURPOSE OF THIS STUDY

Despite this clear charge to develop a monitoring capability, a fundamental question remained: What is an appropriate monitoring role for the NSF? A number of agencies already collect and report education statistics. The Center for Education Statistics within the ED has a long history of collecting data on schooling from a national perspective, and most of the states have assessment divisions within their education agencies that collect comprehensive statistics (Burstein et al., 1985). Most school districts also collect their own achievement test

data. The question of the NSF's role therefore required further exploration, which RAND has attempted to provide.

STUDY METHODS

The process we used to address the question of the NFS's role is outlined in Fig. 3 and summarized below. We first performed an extensive review of the social indicator and educational research literature to develop a basis for formulating a model of the education system and to identify potential indicators for measuring each component of the model. We then applied a set of substantive and technical criteria to the 104 indicators we had identified. Approximately 40 of the measures met our criteria and were used to form the core of a mathematics and science education indicator system.

Once we had created a model of the education system, identified criteria for selecting indicators, and obtained input from the project's national advisory board, we could identify alternative data collection strategies that either exist or could be used to build an indicator system. Interviews with approximately 20 policymakers provided data about the appropriate uses and audiences for a national indicator system.

During this developmental stage, we also surveyed existing databases to determine what information was already available, and we identified areas where new indicators were needed. In addition, we developed a cost model that permitted us to estimate costs for implementing alternative indicator systems. (The model is described in Appendix B.) Using this information, we were able to generate alternative options, assess their likely utility, and provide cost estimates for each.

We identified five generic indicator system options for the NSF:

1. *Status quo.* The NSF can continue business as usual, relying on whatever data are available at the time a report is produced or a policy issue considered.
2. *Patchwork.* The NSF can piece together indicators from available data sources, guided by an integrated and conceptually sound model of schooling.
3. *Cyclical studies.* The NSF can commission indicator studies on a regular basis to provide a time series on such important schooling features as achievement, teacher quality, and curriculum quality. For example, the NSF has twice commissioned national surveys of mathematics and science teachers and principals. Under this option, the NSF would allocate resources for a teacher survey every four years.

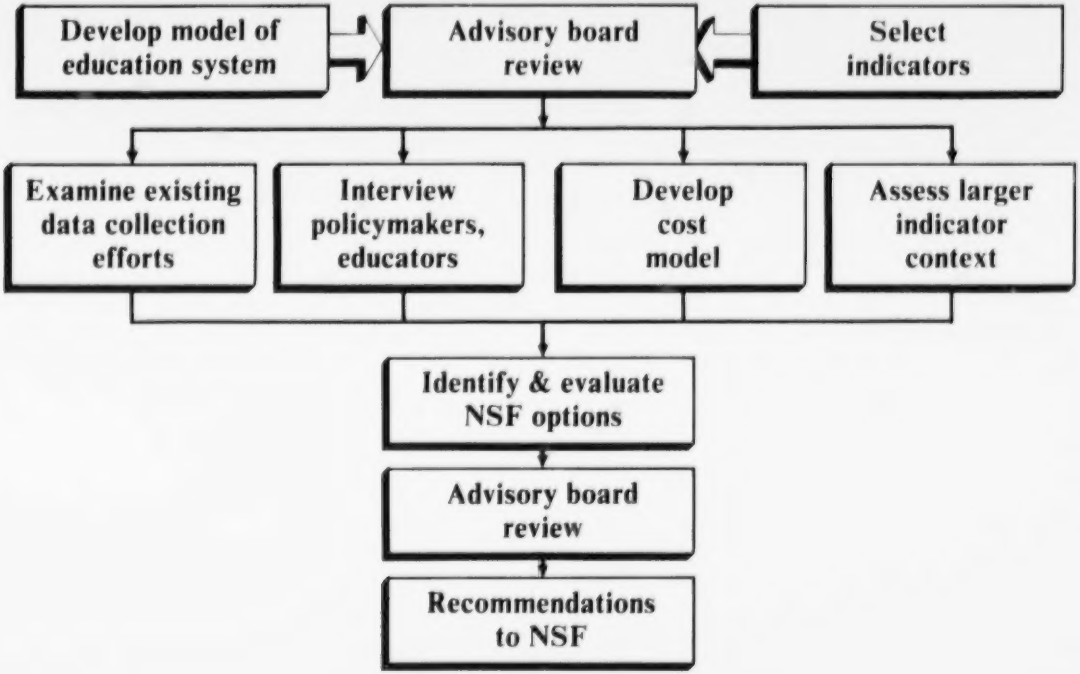


Fig. 3—Process for generating NSF options

4. *Piggyback.* The NSF can collaborate with ongoing data collection programs to obtain more complete data about mathematics and science education.
5. *Independent.* The NSF can develop and field a comprehensive data collection system of its own that spans the major components of education.

Regardless of the option chosen, the NSF should conduct technical studies on indicator development and in-depth policy studies whenever the indicator system identifies potentially important changes.

We evaluated each of the five options according to their utility (i.e., the functions various indicator systems might serve) and their feasibility (i.e., the technical capacity, political support, and costs that the options entail). Finally, we placed monitoring in the context of the NSF's larger agenda, investigated the nature of the decision tradeoffs the Foundation must weigh, and considered the magnitude of the commitment, beyond collecting indicators, that monitoring implies.

This report details our findings at each step of the process and presents our recommendations to the NSF for further work.

II. A COMPREHENSIVE MODEL OF THE PRECOLLEGE EDUCATIONAL SYSTEM

INDICATORS AND INDICATOR SYSTEMS

The notion of indicators is fundamental to our conception of monitoring. Education indicators are single or composite statistics that reflect important aspects of the education system (as economic indicators reflect aspects of the economy). They are expected to tell a great deal about the entire system by reporting the condition of particularly significant features of it. Consequently, educational indicators are the basic unit for monitoring the progress of our complex schooling system.

Whether it is a single or a composite statistic, an educational indicator must enable us to observe and monitor important conditions and trends in the educational system. It should provide insight into the "health," quality, or effectiveness of the system; and it should be useful in the educational policy context. This means that it should provide at least one of the following types of information:

- Information about the educational system's *performance* with respect to attaining desired educational conditions and goals, for example, its progress toward higher achievement, increased student participation, positive attitudes, and equal opportunity; these aspects can be used as benchmarks for measuring progress.
- Information about *central features* of mathematics and science education, particularly those features that enable schools to progress toward desired educational conditions and outcomes—the resources available to education and the process by which students learn. Indicators that provide such information may be seen as bellweathers or leading indicators, i.e., they have predictive value. When these move, other changes can be expected to follow.

Like economic indicators, education indicators have meaning by comparison. When one compares an indicator with itself, one creates a time series and that gives some notion of what is happening. Indicators can be also interpreted by comparison with a normative standard. For example, the National Commission on Excellence in Education in 1983 set graduation standards for high school graduates, and one can compare course-taking with those standards. Alternatively, an

indicator can be compared with the same indicator in another institution or country (e.g., data on academic achievement show that the United States lags far behind Japan and most European countries (McKnight et al., 1987)).

An indicator *system* is a set of interrelated indicators based on a conceptual model. A system is needed because, regardless of whether indicators are single or composite statistics, a single indicator is rarely able to provide information about a phenomenon as complex as education. An indicator system can provide more—and more accurate—information. Ideally, an indicator system measures distinct components of an education system (see Fig. 4) but also provides information about how the individual components work together to produce the overall effect. For example, more than 70 explanations have been proposed for the observed decline in SAT scores shown in Fig. 2 (Wharton, 1977). Was the decline due to changes in teacher quality? An outdated curriculum? A demographic shift? Decreased participation in academic courses? A birth-order effect? In the absence of an indicator system that relates outcomes to student and teacher characteristics, attributes of the curriculum, and student participation, the possible explanations are almost limitless. However, an indicator system can eliminate many possible explanations, leaving the most plausible to be examined in greater detail.

A COMPREHENSIVE MODEL OF SCHOOLING

An indicator system for monitoring a complex system must be based on some kind of conceptual model of how the system works. That model identifies the features that will serve as key indicators of the entire system. The model is often only implicit, but for this study, we decided that the selection of indicators must be based on an explicit model of the precollege science and mathematics education system.

We based our model of the education system on an extensive review of the social indicator and educational research literature, using the information we obtained to select the indicators for measuring each component.[1] The model contains inputs (the human and financial resources available to education), processes (what is taught and how it is taught), and outputs (the consequences of schooling on students from different backgrounds). We also suggested how these elements are likely to be logically or empirically related (Fig. 4). The relationships depicted in this figure, of course, do not constitute a model in

[1]Similar models have been developed to depict the schooling process and to identify and interrelate indicators (Barr and Dreeben, 1983; Hall et al., 1985).

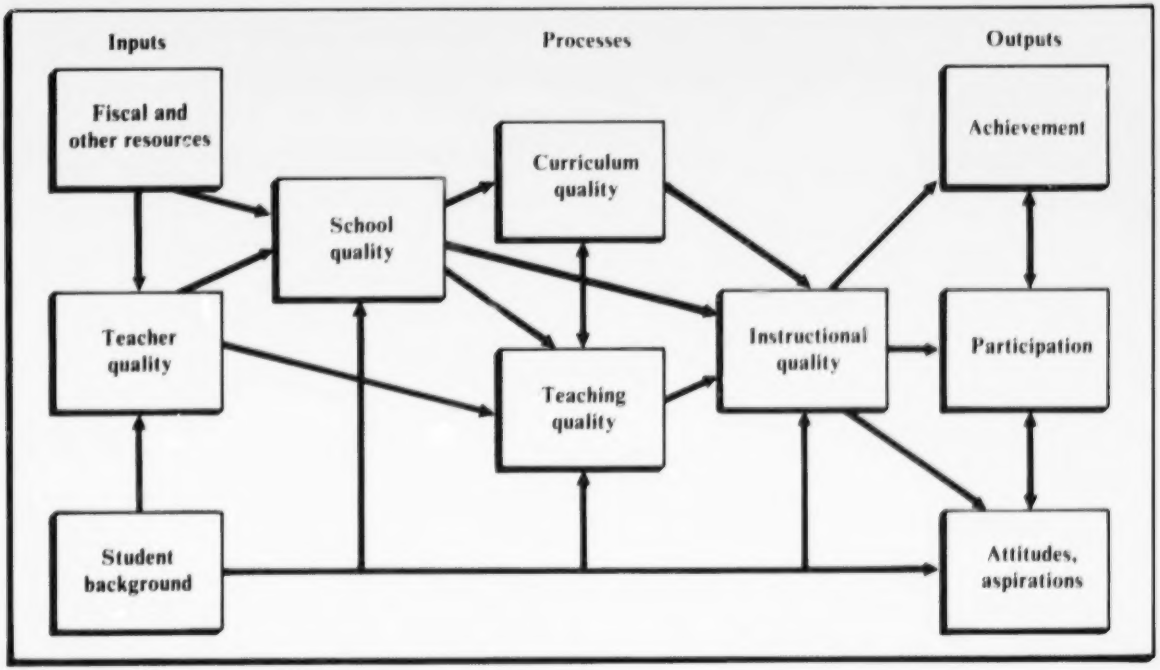


Fig. 4—Linking elements of the educational system

either a strict predictive or causal sense. However, they can serve as a framework, showing logical linkages among elements of the schooling system.[2] Moreover, considerable correlational research supports the links among components. The model also can produce only a static snapshot of the very dynamic process of schooling. Therefore, time-series data must be generated, within the framework provided by the model, that capture change over time.

This comprehensive model provided our conceptual basis for identifying potential indicators.[3] Inputs to the education system include fiscal, capital, and human resources, student characteristics, and teacher qualifications, such as academic credentials. Educational processes may be thought of as a set of nested systems (Barr and Dreeben, 1983). The school translates resources into education; it creates an academic ethos that establishes achievement expectations, and it sets goals and policies so that these expectations can be realized. Curriculum, nested within the school, is the content of education, the medium of exchange between teacher and student. Teachers, operating within the curriculum, draw on their subject matter and pedagogical knowledge to translate the curriculum for students in a comprehensible way. Teachers draw on multiple instructional methods, and this affects outputs such as achievement and course-taking or participation, as well as attitudes about mathematics and science, future aspirations, and the decision to remain in school.

Additionally, we need indicators of the most critical dimensions within each component. For each component, we identified a large number of variables that appear to be important enabling conditions or to have a direct link to the desired outcomes of mathematics and science education. For example, potential indicators of teacher quality would include:

---

[2]One caveat should be noted about our use of the word *indicator*. In a few domains of schooling (e.g., resources), statistical indicators (e.g., per-pupil expenditures, teacher-pupil ratios) have been developed and reported for some time. In these areas, we can more accurately discuss specific indicators. In other domains, where aggregate statistics have not typically been produced and reported as indicators (e.g., schooling processes), we are limited to discussing features (or variables) that are good candidates for indicator development.

[3]A companion volume to this report reviews, in much greater detail, past studies of the major components of schooling. Research on the resources available to education, school quality, teachers, classroom processes, and schooling outcomes was examined to identify the most robust relationships among these components and to determine which measures should be included in an indicator system documenting the status of mathematics and science education. This volume of background papers will be available in early 1988.

- Age and gender
- Personal and career histories
- Attrition rates
- Educational backgrounds
- Pedagogical skills
- Course assignments
- Verbal ability
- Years of experience
- Recency of educational enrichment
- Interest in and comfort with mathematics and science
- Flexibility

The full list of variables and potential indicators is given in Appendix A. The following sections briefly outline the rationale for including indicators from various domains of the comprehensive model, and they suggest the most promising candidates for indicator development.

COMPONENTS OF SCHOOLING AND KEY INDICATORS

Resources and Commitment to Education[4]

In general, research shows that the link between resources (per-pupil expenditure, curricular materials, facilities) and student achievement is tenuous at best (Averch et al., 1972; Coleman et al., 1966; Hanushek, 1981). Nevertheless, resource levels provide basic information about what schools have available to them. On the most basic level, we need to know what financial support schools are provided. At a second level, information is needed about how educational dollars are spent—e.g., how much is allocated for teaching and administrative staff, facilities, equipment, staff development, and curriculum materials. These resource levels provide the parameters in which schools operate; they define the outer limits of what is possible. Whether or not a school provides the equipment and materials necessary for essential instructional activities conveys important information about its potential for educational quality. And when resource levels are coupled with measures of resource use, their power to indicate school quality may increase dramatically (Bridge, Judd, and Moock, 1979; Centra and Potter, 1980; Glasman and Bimaninov, 1981).

Resource levels are also important because they are salient to the public. They are frequently used as informal indicators of school

[4]Although we describe the major components of schooling separately, many are interrelated and some may also be interchangeable. For example, higher-quality teachers may partially compensate for inadequate curriculum materials or vice versa.

quality, both in community discussions and in the popular media. Resource levels thus suggest a level of public commitment to schooling. School or system performance differences across time and settings may also be tempered by available resource levels and community commitment. In short, resource and commitment indicators are important controls for context in any educational monitoring scheme.[5]

School Quality

Indicators of school quality are important because the schools are the places in which policies become programs and activities for children. Decisions about school structure, activities, and behaviors set the conditions for teaching and learning in classrooms. Who the students are; how they are grouped for instruction; the size of classes; the materials and resources available for use; the courses and subjects offered and emphasized; and the overall standards for achievement are just some of the issues determined at the school level.

School-level resources that are central to high quality mathematics and science instruction are well-qualified teachers; class sizes small enough to permit intensive engagement in "hands-on" work; up-to-date textbooks and curriculum materials; laboratories equipped and supplied for hands-on activities and experimentation in the mathematical, biological, physical, and computer sciences; paid assistants to manage and maintain laboratories; and monies for field experiences. These resources signal a school's potential to offer quality education in mathematics and science.

School organizational characteristics that have been shown by research to be central to both instructional quality and student outcomes in mathematics are the amount of time allocated to instruction (Peng, Owings, and Fetters, 1982; Wiley and Harnischfeger, 1974); the curricular emphasis at the school (the teaching resources, the number of courses offered, and the instructional time spent in various subjects) (Borg, 1980; Goodlad, 1984); and the curricular differentiation, i.e., grouping practices and course-taking patterns for various groups of students (Alexander and McDill, 1976; Oakes, 1985; Rosenbaum, 1976).

Other features related to the ethos of the school are also critical. The academic orientation of a school has important links with student outcomes. High schools where faculty and students stress academic accomplishment and student intellectualism foster higher academic

[5]Despite the importance of resources in shaping the conditions under which mathematics and science education occurs, we know very little about the type and level of resources available specifically to mathematics and science. Consequently, this is an area requiring further indicator development work.

achievement and encourage more students to make college plans (Madaus et al., 1979; McDill and Rigsby, 1973). Elementary schools in which students achieve well are characterized by teacher expectations for academic success, student support for academics, a system for monitoring and rewarding academics, greater time focused on learning, and protection of classroom time for teaching and learning (Brookover et al., 1979; Clark, Lotto, and Astuto, 1984).

Finally, one of the most consistent findings is that children learn more when schools involve the parents in instruction—either at home or at school (Epstein and Becker, 1982; Walberg, 1984; Barth, 1979; McDill and Rigsby, 1973). Encouragement from parents may also counterbalance some of the usual achievement barriers faced by poor and ethnic minority secondary students (Rock et al., 1985).

Teacher and Teaching Quality

At the heart of the educational system is the classroom, in which students and teachers interact to produce learning. The quality of this interaction ultimately determines the quality of the educational system. And the quality of teacher-student interaction is greatly affected by the qualities (qualifications, attitudes, training, beliefs) of the teachers assigned to students. But what do we really (i.e., empirically) know about teacher quality? We know that certain features of teachers and teaching are consistently related to student achievement, and that some teacher characteristics and behaviors are more "effective" in some educational contexts than in others. What we do not know is the distribution of these teacher qualities across the teaching force, how this distribution changes over time, or how teacher qualities relate to teachers' practices, working conditions, or career decisions. This is the information that indicators can begin to provide.

Our analysis of the literature on teacher and teaching quality suggests that three types of indicators are needed: (1) teacher characteristics that reflect qualifications, experience, and attitudes toward mathematics and science teaching; (2) descriptors of teaching assignments and conditions; and (3) factors influencing teacher supply and demand, including retention of current teachers and trends in the supply of prospective teachers. Trends in teacher supply are included because there is strong and growing evidence of teacher shortages in mathematics and science education, and because such shortages have important effects on the quality of education students receive (Association for School, College, and University Staffing, 1984; Darling-Hammond, 1984; Howe and Gerlovich, 1982; Johnston and Aldridge, 1984; Rumberger, 1985).

In addition to providing basic indexes of the teaching force, teacher quality indicators should permit educators and policymakers to explore several major issues. Better indicators are needed of how background features (e.g., certification status and levels of college coursework) are related to each other and to other teaching variables (e.g., teacher experience, types of students taught). Because there is so little current agreement about what constitutes qualification to teach (and because certification indicates very different levels of qualification across states), it is important to develop multiple indicators of teacher quality and to relate them to each other. It is also important to examine how—according to these various measures—teacher quality is distributed across classes and students of different types (Pascal, 1987).

Teaching assignments and conditions mediate the influence of teacher qualifications on instructional quality and performance (e.g., Mera et al., 1983; Welch, 1984). Well-qualified teachers, according to any given indicator, may perform less well if they are assigned to teach out of their field, have extraordinarily large class sizes, or lack the necessary material and equipment. Thus, indicators of teaching quality should include not only the conditions under which mathematics and science teachers work, but how these conditions are distributed across teachers and schools of different types, and how they influence teaching practices.

Many working conditions also affect mathematics and science teachers' probability of remaining in teaching, through their effects on teacher satisfaction and commitment (Locke, 1976; Lortie, 1975; Rosenholtz, 1985; Rosenholtz et al., 1985). In addition to class size and course load, other working-condition indicators likely to affect teachers' attitudes and retention are teachers' satisfaction with materials and equipment (of greater importance to science teachers, perhaps, than to any other subject-area instructors (see Weiss, 1978; Welch, 1984)), salary level, and autonomy and collegiality—two variables that contribute to teachers' feelings of empowerment and efficacy (Rosenholtz, 1985; Rosenholtz et al., 1985).

Finally, indicators of class size, course load, misassignment, and teacher qualifications should be examined separately for new hires. The assignments and qualifications of new hires, especially, can suggest the influence of current supply and demand conditions on hiring and assignment decisions. Changes in policy decisions related to teacher training and certification are also revealed most clearly in the qualifications of newly hired mathematics and science teachers.

Classroom Processes: Curriculum and Instruction

Curriculum is usually defined as content—the topics, concepts, processes, and skills that students are taught in science and mathematics classes. Curriculum content is the "medium of exchange" in school. The content taught in school clearly makes a difference in what students learn. Students are more likely to learn particular mathematics and science topics, concepts, processes, and skills when teachers teach them (Crosswhite et al., 1984; Husen, 1967; Wolf, 1977).

But classroom processes include more than simply the content that is covered. They include the depth to which the content is explored; the way it is sequenced; the textbooks and materials used; and the mode of presentation teachers use. The classroom is a place where policy is often substantially modified. For example, teachers' decisions regarding what topics to cover, grading policy, and methods of presenting material can influence the degree to which policymakers' intentions are implemented. The teacher is the one who makes decisions regarding goals of the class (Shavelson, 1983), the inclusion of or exclusion of topics in the syllabus (Schwille et al., 1981), whether students have learned enough to progress to new topics (Barr, 1980), methods of presenting topics, grouping practices (Webb, 1980), standards for evaluating pupil progress, and task structures (Doyle, 1977). Perhaps most important, classroom processes operationalize the substantive goals teachers intend to accomplish with children (e.g., to have students acquire basic facts and skills, or to have students develop problem-solving strategies).

Indicators must be developed that *describe* these domains of mathematics and science curriculum and instruction and allow policymakers and educators to make judgments about their quality—judgments based on comparisons among states and localities, comparisons over time, and comparisons with curriculum standards set by subject-area experts (e.g., the Conference Board of Mathematical Sciences). Additionally, curriculum indicators should enable analysts to track the effects of various policies on the curriculum.

Classroom-policy and instructional-practice indicators will provide information about what students actually experience in classrooms. Indicators of these central curricular and instructional features are necessary for an accurate picture of the science and mathematics education system. They will describe what students have an opportunity to learn—a critical factor in what they actually do learn (Berliner, 1979; Fisher et al., 1980).

Essentially, we need indicators of the following classroom policies and practices: (1) how much time teachers spend on science and mathematics instruction—particularly at the elementary level, where evidence indicates that teachers vary widely in the time they devote to these subjects (Raizen and Jones, 1985); (2) the breadth and depth of specific concepts, topics, and skills to be taught; (3) sequencing and pacing of lessons and courses; (4) the mode of presenting science and mathematics to students, including the task structures teachers use (e.g., whether content is presented as information in lectures and textbooks or as problems to be solved in field- or laboratory-based experiences); (5) the textbooks and materials used; (6) the goals and objectives of science and mathematics lessons and courses; and (7) grouping practices and accompanying differentiated lessons and courses for students with different abilities or different future plans (e.g., college, work).

Curriculum-materials "quality" indicators are needed to provide information about the quality of the curriculum presented in state, district, and school curriculum guides, and in the most commonly used textbooks and materials. These features also set the boundaries of students' learning opportunities and outcomes. Two indicators will be useful: (1) the congruence of the science and mathematics materials actually being used in classrooms, with "expert" judgments of the ideal curriculum; (2) the pedagogical appropriateness of the materials (e.g., how well the texts and guidelines match the cognitive needs of students).

Individual-Level Outcomes

The domain of individual outcomes includes (1) students' *achievement*, i.e., knowledge, understanding, and use of concepts and skills in mathematics and science; (2) students' *participation* in mathematical and scientific activities within and outside of school; and (3) students' *attitudes* toward and self-confidence about these subjects.[6]

Achievement. Achievement outcomes are important because (1) passing or failing a course has a dramatic effect on further participation in related coursework; (2) even those students who pass a course might be in a disadvantageous position if they do not fully understand certain concepts; and (3) achievement is equated by policymakers, politicians, educators, and the public with the effectiveness of the

[6]In selecting outcome factors, we concentrated on those proximate to precollege mathematics and science education. Therefore, our analysis does not include such factors as post-secondary educational choices or employment opportunities. An expanded indicator system could, of course, incorporate them.

education system. Most important, achievement outcomes should be given the highest priority in an indicator system because they represent the essence of schooling: whether students have learned what they have been taught.

Outcome indicators should show the extent to which students are learning problem-solving skills and developing conceptual understanding ("higher-order thinking skills") in mathematics and science. They should also suggest whether students are developing attitudes that foster achievement and continuing interest in mathematics and science, and they should permit analysts to explore relationships among achievement, attitudes toward mathematics and science, self-confidence, and expectations that those subjects will be relevant to students' future careers. Such analyses will contribute to a better understanding of which attitudes exert the most influence on achievement and continuing participation in mathematics and science.

Also needed are composites of outcome variables that can be used to communicate the multidimensional nature of educational outcomes more effectively. Such composites can indicate relationships that may contribute to our understanding of policy tradeoffs or patterns of outcomes that occur in schools (cf. Klitgaard, 1974). For example, if good composite indicators were available, we might be better able to measure the tradeoffs involved in requiring more academic courses for high school graduation (e.g., exposing students to more content vs. the effects on lower-achieving students of high course-failure rates and reduced curriculum choice).

Finally, a variety of composite indicators are needed to address questions about mathematics and science outcomes. For example, multidimensional indicators might explore patterns of strength and weakness in achievement among various topics—these indicators could pinpoint areas where the NSF should focus attention on curriculum improvement. A second composite indicator might contrast student memory for factual knowledge with achievement in understanding and problem solving.

Achievement-test scores provide the most readily available indicators of achievement. However, none of the existing tests actually measure some of the most important dimensions of achievement. To assess achievement, tests need to measure students' substantive knowledge, their conceptual understanding, and their skill in applying them to solve both academic and "everyday" problems. To measure understanding and problem-solving skills validly, tests should be sensitive to (1) the curriculum students have been exposed to, (2) differences in student aptitudes, (3) students' conceptual misunderstanding, and (4) how students apply concepts and skills deemed important in

scientific and mathematical enterprises. Our analysis shows that the existing tests do not meet these criteria. In the short run, an indicator system could use a specific combination of data from various achievement tests. However, new methods should be devised to assess understanding and problem solving more validly. The importance of this recommendation goes beyond testing: Achievement tests exert a profound influence on what is taught in classrooms.

Participation. Participation in mathematics and science is a precondition for achievement; it provides a concrete, albeit imperfect, indicator of whether students are likely to progress toward becoming mathematicians, scientists, and engineers; and it provides a useful, indirect indicator of student attitudes.

To assess participation, indicators should reflect the taking of required and elective mathematics and science courses and the topics and skills covered. These two pieces of information can predict achievement and students' decisions as to whether or not to continue into technologically related employment. Indicators should also show the degree to which students participate in extracurricular activities related to mathematics and science, and students' intended college majors or career choices. Extracurricular participation predicts achievement (Shavelson, 1985), is related to education/career decisions (Berryman, 1983; Munday and Davis, 1974), and may provide insights into the scientific manpower pipeline.

Attitudes. We assign a lower priority to measures of student attitudes, because attitudes are not as highly valued by policymakers and others as are achievement and participation. Furthermore, there are difficulties in measuring and interpreting correlates of participation and achievement, such as scientific attitudes, favorable attitudes toward mathematics and science, and self-confidence in one's abilities (Munby, 1983; Welch, 1983). Nevertheless, we conclude that attitudes should be included in an indicator system, because they are perceived by many to be a relevant outcome of schools, they are thought to provide an early sign of mathematics and science dropouts, and data on them are easy to collect.

One possible attitude indicator is the anticipated usefulness of mathematics and science for one's adult life. Berryman (1983) has found that girls who are interested in mathematics and science careers are more likely to take classes in these subjects. The difficulty with this indicator is that, like enjoyment, the perceived usefulness of mathematics and science cannot be interpreted as solely related to the school—social desirability and societal factors clearly play a large part in students' responses to this type of question.

Participation and Outcomes of Special Populations

The educational status of women, the poor, and non-Asian minorities in mathematics and science warrants the development of indicators to monitor both educational opportunities and achievements of these groups. Their typically lower levels of achievement and participation are cause for concern about the impact on the national welfare of this increasing segment of the future workforce. Underrepresentation and lower achievement in science and mathematics also relate to the long-standing federal responsibility for ensuring equal educational opportunity for minority and disadvantaged populations.

Data describing how the precollege mathematics and science opportunities and experiences of women, minorities, and the poor typically diverge from those of more successful student groups (Asians and white males) provide considerable evidence that schooling factors may contribute to unequal outcomes (Berryman, 1983; Levin, 1972). Patterns of schooling differences accompany those of lower achievement and underparticipation. Both patterns begin early in the educational process. Race-, class-, and gender-related differences appear to result, at least in part, from the insufficient and unequal access of women, minorities, and poor students to those school conditions that work in favor of high achievement, continued participation, and positive attitudes.

Equity indicators would monitor the distribution of schooling conditions that are factors in the participation of special populations. Specifically, indicators should include measures of resources and opportunities available at schools of different types, i.e., those serving different student populations. Such indicators could focus specifically on the distribution of science and mathematics resources, teacher quality, and actual schooling experiences. Indicators also should assess the extent to which resources and opportunities are available to different groups of students *within the same schools*. Incorporating these indicators into a national educational monitoring system would permit tracking progress toward improved outcomes for these "at risk" students. They would also chart efforts toward providing more equitable opportunities for them in science and mathematics.

The Policy Context

Although it is not explicitly depicted in Fig. 4, the larger policy environment in which schooling occurs profoundly influences education in a variety of ways. Federal, state, and local policies largely determine the level and type of resources available to education. Although their

effects on other components of schooling are typically less direct, these policies can also influence who is allowed to teach, what content is taught, and even how it is taught. The priorities of the policy community may also signal to educators which outcome factors are most important at a given time and, thus, the criteria on which those educators will be held accountable. What is perhaps most striking about the education reform policies of the last four years is that they have moved the influence of policy further down into the educational system than has traditionally been the case. In the past, policy typically focused on how schools were financed and governed; now its scope also includes what is taught and who teaches it. Consequently, the larger policy context must be taken into consideration in attempting to explain changes in the major components of schooling.

CONCLUSIONS: USING A COMPREHENSIVE MODEL FOR IDENTIFYING INDICATORS

Current indicator systems—to the extent that such systems exist—tend to consist of rather disjointed pieces of information about such topics as fiscal resources, teacher characteristics, and student test scores. Although a growing body of research demonstrates that many of these categories are correlated, if not causally linked, few producers of education statistics have used these research findings as the basis for a more systematic interpretation of educational trends. The comprehensive model of the mathematics and science education system presented here permits indicators about student performance to be placed in a valid context that controls not only for fiscal and human resource inputs to elementary and secondary education, but also for the nature of the schooling process itself. Furthermore, it ensures that the choice of indicators moves beyond conventional wisdom about what makes a difference in schooling to decisions based on the most robust research findings.

However, the design of a good indicator system requires more than technically valid indicators. The measures selected must also be useful to their primary audience, the policymakers responsible for the conduct of mathematics and science education. Therefore, we next attempt to assess the utility of the 104 indicators that our research has identified as central to a comprehensive model of schooling.

III. DEVELOPING A USEFUL INDICATOR SYSTEM

The quality and credibility of an educational indicator system depend on technical factors related to how schooling domains are conceptualized and then measured. Based on our review of the education research literature and continuing consultation with the project's national advisory board, we believe that the indicators we have chosen validly reflect the central features of mathematics and science education.

Yet we know from the history of earlier efforts to identify social indicators that technical quality is not sufficient to guarantee the continuation of such systems (de Neufville, 1975; MacRae, 1985). The systems must also produce information useful to the policy community if they are to survive as publicly supported endeavors.

Usefulness to the policy community depends on much more than just the dissemination of indicator data. It also hinges on factors such as the indicators contained in a system, how they are conceptualized and measured, the level at which they can be disaggregated, and the way they are analyzed and reported.

The social indicators movement that burgeoned in the 1960s and then quickly died contained an important lesson about the relationship between policymakers' needs and the process of indicator development. The social indicator producers of the 1960s failed to "deal with the style, objectives and constraints of decision-makers," and information users failed in "not establishing the systems which would permit them to choose the information they needed to make use of it" (de Neufville, 1975). The lesson from this experience is that education indicators must be developed iteratively with decisionmakers to ensure that the information produced meets their needs.

Therefore, before developing criteria for determining which indicators would be most useful to policymakers, we interviewed approximately twenty NSB members, NSF managers, Congressional staff, Office of Management and Budget (OMB) and Executive Office staff, representatives of state policymakers, and NSF science and engineering education constituents. These interviews are summarized below. (The interview guide is reproduced in Appendix D.)

POLICYMAKER INTERVIEWS

Respondents' Information Needs

Our respondents were in general agreement about the kind of data they need. Their information categories correspond roughly with many of the indicator domains included in our comprehensive model of schooling—i.e., student achievement, teacher qualifications, course-taking and course content, resources available for mathematics and science, and students' laboratory experiences. School climate variables were rarely mentioned. Most respondents indicated that they wanted data disaggregated by sex and ethnicity.

Concern about sex and ethnicity was further emphasized by the significant number of respondents who discussed pipeline issues and the need to monitor the number and type of students who are preparing for mathematics and science careers. These respondents argued that such tracking needed to focus on students early in their school careers (i.e., as early as elementary school) and should provide the kind of information that would allow agencies such as the NSF to increase the successful participation of women and minorities in science and mathematics.

There were some differences across respondent role positions in their views of the relative priorities and amount of information needed. For example, as would be expected, those furthest from the science and mathematics educational enterprise (e.g., the White House Office of Science and Technology Policy (OSTP) and governors' aides) have little need for curriculum indicators, beyond course requirements and enrollments, while NSF staff members have the most detailed information needs.

The Uses of a Mathematics and Science Monitoring System

An indicator system could conceivably be used for any or all of the following purposes:

- To provide a broad overview of the status of mathematics and science education in the United States.
- To serve as an accountability mechanism by reporting data about how well schools and students are performing.
- To improve policy and practice by providing information about which approaches seem to be working.

Clearly, neither the NSF nor designers of any other indicator system can control the uses to which the information generated is put. However, they can design the system in ways that minimize the likelihood that data will be used for purposes they deem inappropriate (e.g, by the

selection of indicators, the level at which they are disaggregated, and how they are reported).

With only one exception, respondents said that describing the condition of mathematics and science education is the most important function a monitoring system can serve. However, their notions of what should be included in this descriptive or overview function cover a broad spectrum. Respondents talked about wanting to know what the "health" of mathematics and science education is, and about needing a system that identifies trends and aids in anticipating future problems. In addition to an interest in general trend data, respondents were particularly interested in information about those factors over which they can exert some leverage. Within this general framework, they want data that will help inform choices they need to make, given limited resources. One such example is the need to understand the tradeoffs involved in spending money on teacher salaries, as compared with purchasing laboratory equipment or reducing class size.

Our respondents uniformly argued that a national monitoring system should not be used to promote accountability for mathematics and science education. In their view, accountability, beyond that needed to make certain that federal programs are operating appropriately, is not a federal responsibility. At the same time, respondents believe that the federal government can provide information that states might use as part of their own accountability systems.

Respondents also felt that since improving practice is clearly within the mission of the NSF and other federal agencies such as the ED, this function is appropriate for a national indicator system. For example, respondents argued that in its leadership role in mathematics and science, the NSF should provide information about exemplary practices—those that are consistently related to valued outcomes for students. Respondents viewed the improvement of policy and practice as an appropriate purpose for a national indicator system.

However, some respondents noted that identifying exemplary practices requires a different type of information than can be generated from a basic monitoring system. Several respondents also suggested that more in-depth studies or special analyses, linked to the larger indicator system, might be a way to deal with this problem.

Indicator System Audiences

Respondents identified several kinds of audiences. The first is the NSF, which could use indicator data for internal planning purposes. Other federal agencies or institutions such as Congress were rarely mentioned as potential audiences. The most commonly mentioned

audience, along with the NSF, consisted of state policymakers, particularly governors and legislators. A number of respondents argued that if changes were to be made in mathematics and science education, these were the people who would have to make them. The emphasis on state policymakers as an audience was reinforced by respondents' beliefs that indicator system data should be disaggregated to the level of individual states. Local school districts were also mentioned, but with less prominence.

Several respondents mentioned higher education as another indicator system audience, because policymakers at universities and colleges need to know what kind of students they will be receiving in order to plan the type of education to offer them. The business community, particularly high-technology industries, was also mentioned; with the right information, industry can be an advocate for the needs of mathematics and science.

Coordination with Other Indicator Efforts

All respondents agreed, in theory, that the NSF should piggyback its indicator efforts on existing databases. The most frequently mentioned database was the National Assessment of Educational Progress (NAEP). A number of respondents argued that NSF participation in the NAEP would allow mathematics and science performance data to be collected more frequently and in greater depth. In addition, they argued that such sharing would reduce the data collection and time burden on local districts and schools.

However, several NSF constituents expressed concern about the objectivity and reliability of the ED's data collection activities. Similarly, senior NSF staff alluded to past problems that the NSF had encountered in collaborative efforts with the ED. Several Congressional and OMB staff members also suggested that NSF and ED interests may not always be consistent. Even several who argued that it would be bad public policy for the NSF not to add funds to the NAEP noted that ED data collection activities have been ad hoc in the past.

The concerns expressed by respondents raise several issues related to NSF collaboration with the ED: (1) consistency of interest between the two agencies; (2) the transaction costs associated with interagency agreements; (3) the stability of ED priorities; and (4) the quality of ED data collection efforts. Joint participation in indicator efforts requires overlapping interests between the two agencies. By virtue of its organizational mission, the ED must deal with education generally, in contrast with the NSF's more focused interest in mathematics and science

education. As long as interests are reasonably consistent and funding is the only question, however, both the broader and more specialized missions can be accommodated. Problems may arise at the operational level when the ED wants to cover a broad range of topics and the NSF wants to concentrate on a few mathematics and science-related subjects. Given the very real opportunity costs involved in data collection within strict time limits, these different interests could make cooperation more difficult.

Another issue relates to the additional transaction costs connected with any interagency agreement. However, over the past few years, the NSF has successfully supplemented funding for several ED data collection efforts (e.g., for an expanded sample of mathematics and science teachers in the National Educational Longitudinal Study of 1988 (NELS '88) and new approaches to the testing of science in the NAEP). The administrative procedures established for these activities may provide a basis for further collaboration.

In questioning the stability of the ED's priorities and the quality of its data, respondents raised issues that the Department has tried to address recently. Its priorities in data collection have not been stable over the past decade. The problem was particularly acute in the early 1980s when its data-gathering arm experienced a significant budget reduction that resulted in the elimination of some data collection activities and decreased sample sizes and frequency of collection (Chelimsky, 1986). Problems with data quality and lack of timeliness are widely known (Levine, 1986; Plisko et al., 1986), and in the redesign of its core data collection activities, the ED has taken steps to resolve them. Clearly, respondents' concerns are well-founded, given past history. However, evidence suggests that with increased funding, the ED may be able to overcome its problems and thus provide a sounder basis for collaboration with the NSF on joint indicator efforts.

An Appropriate Funding Level

With few exceptions, $10 million per year was the outer limit mentioned as the amount the NSF should spend for an indicator system; most respondents mentioned a level closer to $5 million per year.

Implications for Indicator System Options

The interviews suggested some major implications for the indicator options we designed:

- The indicator options selected must provide data on issues related to the scientific manpower pipeline. This concern transcends partisan ideologies, and, in effect, provides a pragmatic, policy-relevant rationale for paying close attention to equity issues.
- Because state policymakers are a critical indicator system audience, the systems developed must allow data disaggregation to the state level.
- Although some strategy for cooperating with other data collection agencies such as the ED might initially appear to be a very attractive option, the feasibility issues inherent in relying on interagency cooperation for data must be considered.
- The $5 million to $10 million per year funding levels suggested by most respondents must be taken seriously in identifying a range of options.

CRITERIA FOR SELECTING INDICATORS

The interview data and ongoing discussions with NSF staff and the project's national advisory board generated eight criteria to be used in selecting elements of the schooling process to incorporate into an indicator system. To be included, an indicator should:

- Provide information that *describes central features of the educational system*—for example, the amount of financial resources available, teachers' work load, and school curriculum offerings. Even though research has not as yet determined the relationship of some of these features to particular outcomes, information is needed about them to understand how the system works and because policymakers and the general public care about factors such as per pupil expenditures and class size.
- Provide information that is *problem-oriented*. Indicators must provide information about current or potential problems—for example, factors linked to teacher supply and demand, or to the changing demographics of urban areas.
- Provide information that is *policy-relevant*. Indicators should describe educational conditions of particular concern to policymakers and amenable to change by policy decisions. For example, indicators of teacher characteristics such as educational background and training are policy-relevant, since they can be changed through legislation or regulations governing teacher licensing.

- *Measure observed behavior rather than perceptions.* Indicators will be more credible if they assess actual behavior rather than participants' opinions or judgments. For example, the academic rigor of schools is better measured by course requirements and offerings than by principal, teacher, and student perceptions.
- *Provide analytical links among important components.* Indicators will be more useful if they permit the relationships among the different domains of schooling to be explored.
- Generate data from measures generally accepted as *valid and reliable.* Indicators should measure what they are intended to measure and should do so consistently.
- Provide information that can be *readily understood by a broad audience.* Indicators need to be easily comprehensible and meaningful to those beyond the immediate mathematics and science community—to policymakers, press, and the general public.
- Be *feasible* in terms of timeliness, cost, and expertise. Indicator data need to be produced within a time frame that is compatible with policymakers' decision cycles and within given cost constraints; they should also be collectable, analyzable, and reportable within current levels of expertise.[1]

Using these criteria, we selected about 40 potential indicators that both reflect the major components of schooling and meet basic standards of usefulness to the policy community. These measures then became the core around which different indicator system options were generated.

Applying these criteria, however, meant that some indicators that parents and others would probably have liked had to be eliminated. For example, one indicator that many people are interested in is the quality of teachers' explanations of scientific concepts to students. Unfortunately, because it cannot currently be measured reliably, this indicator had to be eliminated.

This exercise suggested that some potential indicators, while not sufficiently developed to be included in an indicator system at this time,

[1]Some of these criteria create tradeoffs with each other. For example, describing the central features of schooling assumes that indicators will be relatively stable over time. On the other hand, providing policy-relevant information means that indicators may need to change as areas of policy emphasis change. Because of these competing requirements, indicator systems are likely to expand over time. However, if the system is based on a comprehensive model of schooling, the addition of new indicators can be guided by that model, thus ensuring that the new indicators are conceptually sound. One way of meeting these dual information needs is to develop a system that measures the central features of schooling using a stable set of indicators, but also provides "empty slots" that can be used for policy-specific indicators that change as policies shift.

are nevertheless critical to a better understanding of mathematics and science education and should be part of a developmental research agenda for the NSF. Once these factors can be measured reliably, they can be incorporated into an indicator system. A number of schooling components fall into this category. For example, teacher flexibility in the classroom is known to be related to student achievement. Under current cost constraints and time considerations, such factors cannot be measured adequately, but they comprise a future indicator research agenda for the NSF to consider.

Having selected a set of individual indicators that are both conceptually sound and likely to provide useful information to policymakers, we next combined those indicators into a range of different systems, and then attempted to assess the quality, feasibility, and potential uses of each system.

IV. INDICATOR SYSTEMS: FIVE GENERIC OPTIONS

This section describes five indicator options that reflect both research knowledge about conceptually sound indicators and policymakers' information needs. The options differ in the extent to which the information they generate can answer policy questions relevant to mathematics and science education; in their quality; and in their cost and ease of use. Section V compares their potential uses and feasibility.

The five options are:

- *Status quo.* The NSF would rely on whatever data were available at the time a report is produced or a policy issue is considered.
- *Patchwork.* Indicators would be pieced together from existing data sources, guided by an integrated and conceptually sound model of schooling.
- *Cyclical studies.* The NSF would commission data collection on a regular basis to provide a time series on several domains of schooling such as achievement, teacher quality, and curriculum quality.
- *Piggyback.* The NSF would buy into ongoing data collection programs to obtain more complete data about mathematics and science education and about areas of particular priority.
- *Independent.* The NSF would develop and field a comprehensive data collection system of its own that spans the major components of schooling.

THE STATUS QUO OPTION

Status quo indicators are developed from the data about precollege science and mathematics education that are available at the time the indicators are needed. One set of status quo indicators is currently found in the chapter on "Precollege Science and Mathematics Education" in *Science Indicators*, published by the NSF every two years. For example, indicators for the 1985 edition included currently available statistics on student achievement and aptitude, course-taking, and the supply and demand of teachers and their qualifications. The 1985 edition also included statistics showing international comparisons in

student achievement, using recently analyzed data from the Second International Mathematics Study. The 1987 edition will include statistics on characteristics of school science and mathematics programs and on teacher qualifications, since new data from the 1985 Survey of Science and Mathematics Education will be available.

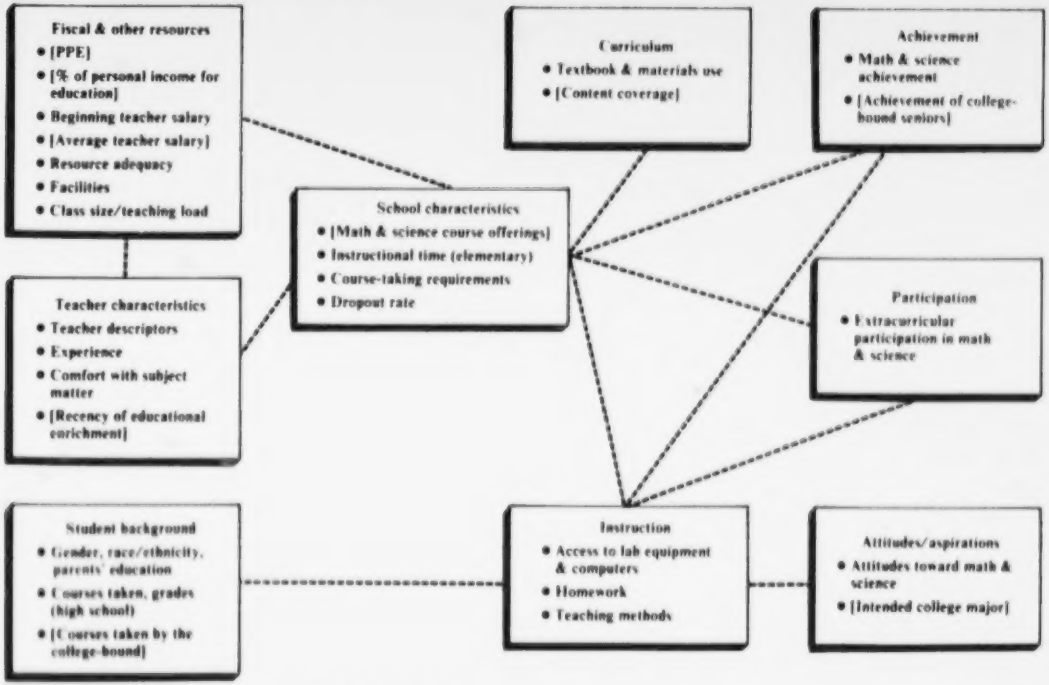
Status quo indicators do not represent an indicator *system*, since they neither stem from an overarching conception of what indicators are needed, nor are they necessarily part of a systematic data collection effort. Essentially, status quo indicators are ad hoc. Their substance, quality, level of aggregation, consistency over time, and, consequently, their usefulness all vary, depending on what data and analyses are available at a given time.

THE PATCHWORK OPTION

Another option for a national mathematics and science indicator system is a patchwork of indicators that combines existing data from a variety of sources. A patchwork indicator system is basically a more institutionalized, better-structured version of the status quo. This option assumes that the NSF would collect no new data, and that only currently available data would be included in the system. The sources of data on critical indicators would be those that provide current national (and state-level, if available) estimates on variables of interest. These sources include ongoing and relatively stable collection efforts, longitudinal studies, and one-time or infrequently repeated surveys.[1]

Figure 5 illustrates what one such system might look like in 1987. Basically, it takes the comprehensive model of schooling outlined in Sec. II and fills it in with data from existing sources that are considered to be the most stable in their collection. (The indicators and data sources are listed in Table 1.) Data sources that are only in the planning stage, such as the Elementary/Secondary Integrated Data System (ESIDS) now being considered by the ED, are not included. Most of the indicators in the patchwork system come from the NAEP because it displays trends in mathematics and science education over time and allows the exploration of possible links among different schooling domains.

[1]Such databases include the NAEP, the International Association for the Evaluation of Educational Achievement (IEA), public/private school surveys, the 1985 National Survey of Science and Mathematics Education (NSSME), the National Longitudinal Study of the High School Class of 1972 (NLS), High School and Beyond (HS&B), the SAT—Mathematics Achievement, and the American College Test (ACT) Mathematics and Science Assessments.



Brackets indicate data available from sources other than NAEP.

Dashed lines indicate that data can be linked analytically.

Fig. 5—An example of a patchwork indicator system

Table 1

PATCHWORK INDICATOR SYSTEM

Indicator	Data Source
Fiscal and other resources	
Per-pupil expenditure	Common Core
Percent of personal income for education	Census
Beginning teacher salary	NAEP
Average teacher salary	NEA
Resource adequacy	NAEP (teacher)
Facilities	NAEP (teacher, school)
Class size/teaching load	NAEP (teacher)
Teacher characteristics	
Teacher descriptors (age, major, minor, race, gender)	NAEP (teacher)
Teacher experience	NAEP (teacher)
Comfort with subject matter	NAEP (teacher)
Recency of educational enrichment	NSSME
Student background	
Race, gender, parents' education	NAEP (student)
Courses taken, grades (high school)	NAEP (student)
Courses taken by college-bound	SAT, ACT
School characteristics	
Math and science course offerings	NSSME, NAEP (school)
Instructional time, math and science (elementary school)	NAEP (school)
Course-taking requirements	NAEP (school)
Dropout rate	NAEP (school)
Curriculum	
Textbook and material use	NSSME, NAEP (teacher)
Content coverage	IEA
Instruction	
Access to lab equipment and computers	NAEP (teacher)
Amount of homework	NAEP (student)
Teaching methods	NAEP (student)
Achievement	
Math and science achievement	NAEP (student)
Achievement of college-bound seniors	SAT, ACT
Participation	
Extracurricular participation in math- and science-related activities	NAEP (student)
Attitudes	
Attitudes toward math and science	NAEP (student)
Intended college major	SAT, ACT

With a patchwork system, the indicator data reported by the NSF would be analyzed within a comprehensive model of schooling, which would guide the choice of indicators. To the extent that appropriate indicators were available, the system would reflect the major components of schooling, including the links among those components. However, the potential for linking data across existing datasets would be limited, given different samples, variable definitions, and time periods. At the same time, this approach is cost-efficient because it does not require mounting new, large-scale data collection efforts. The system would provide a uniform set of analyses that, depending on data availability, could be fairly consistent over time.

THE CYCLICAL STUDIES OPTION

Cyclical studies would increase the depth of information available from any of the other indicator options by concentrating limited financial resources on a few issues of particular interest to the NSF. The data collection effort would be separate from other ongoing indicator projects, enabling the NSF to explore issues related to mathematics and science education in greater depth, without encroaching on other data-gathering institutions' priorities. By commissioning a single study each year, the NSF could keep the costs within limits, while allowing extensive exploration of each topic every four to eight years. For example, a study of teaching quality might be scheduled the first year, a curriculum study the second, educational uses of technology the third year, and achievement the fourth. Then the cycle would begin again, thus creating a time series.

Although cyclical studies could stand alone, ideally, they would complement whichever of the indicator options the NSF might select. Used in this way, cyclical studies could be based on smaller samples and could explore topics in ways that a large-scale data collection effort could not. For example, a cyclical study might assess student achievement with "hands-on" instruments that present students with a set of laboratory apparatus and ask them to perform a scientific task, rather than with the traditional pencil-and-paper test. Similarly, a study of student course-taking could rely on an analysis of student transcripts rather than student self-reports, thereby generating more reliable and comprehensive estimates.

THE PIGGYBACK OPTION

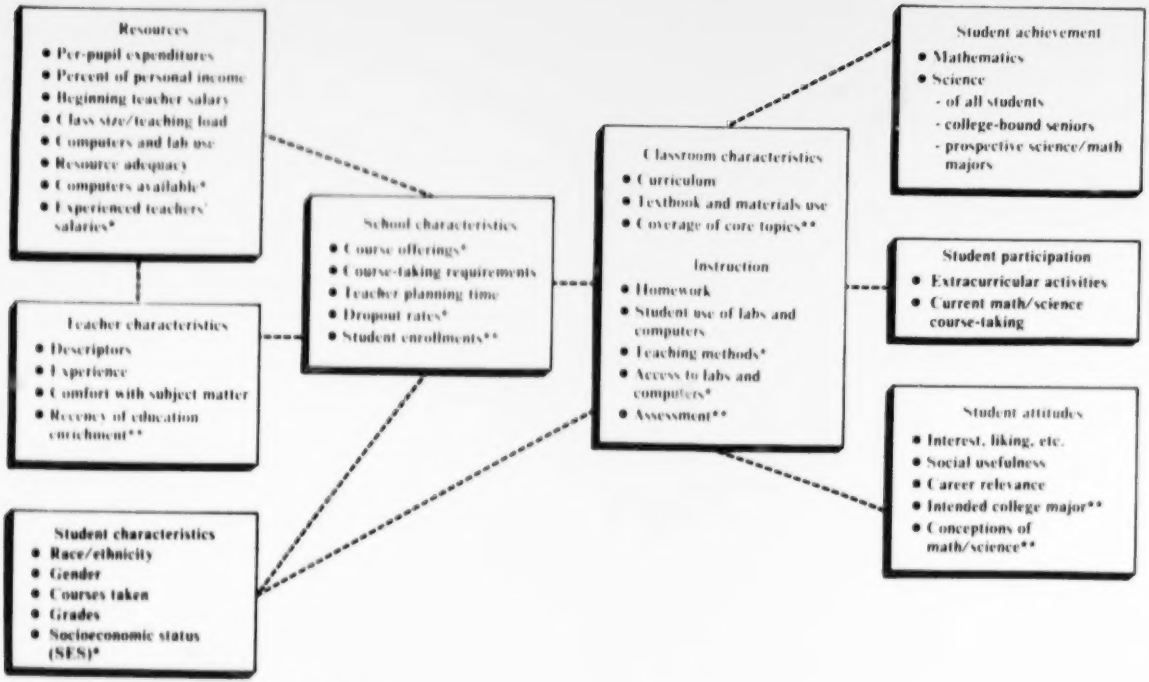
A piggyback indicator system would upgrade the patchwork system with NSF-sponsored expansions of current data collection efforts, primarily through the NAEP. By collaborating with the ED and providing support for additional development and data collection costs, the NSF could design an indicator system that would provide reasonably complete and timely information about the condition of mathematics and science education.

The piggyback system illustrated in Fig. 6, using the variables listed in Table 2, is essentially an expansion of the current NAEP background surveys administered to principals, teachers, and students at schools participating in the mathematics and science assessments. This example is meant only to be illustrative; other systems could conceivably be piggybacked on other data collection efforts. The system in Fig. 6 modestly expands the number of features of the educational system measured in the current NAEP. Its primary contributions result from a changed sampling strategy that would permit linking teacher and student data; expanded sample sizes to permit greater data disaggregation (by state and by social class within ethnic groups on a national basis); and more frequent administrations (e.g., every two years instead of the current four-year cycle) to ensure the inclusion of timely information.

THE INDEPENDENT OPTION

Under an independent option, the NSF would assume total responsibility for a mathematics and science education indicator system, rather than relying on existing or anticipated efforts. The selection of indicators would be guided by the comprehensive model of schooling and the utility criteria listed in Sec. III. Depending on the availability of resources, this system could be quite limited or very comprehensive. Such a system might involve collecting the same data contained in the piggyback option, but without collaboration with other agencies or other data collection efforts. The NSF would assume full responsibility for the design, data collection, analysis, and reporting of mathematics and science education indicators.

Each of these alternatives has distinct advantages and disadvantages. In Sec. V, we assess their usefulness and feasibility and discuss some of the issues they raise.



* Indicates expansion of existing NAEP items.

** Indicates addition of new NAEP items.

Fig. 6—An example of a piggyback indicator system

Table 2

PIGGYBACK INDICATOR SYSTEM

Indicator	Data Source
Fiscal and other resources	
Per-pupil expenditures	Common Core
Percent of personal income expended on education	Common Core
Beginning teacher salary	NAEP (school)
Average teacher salary	NEA
Class size/teaching load	NAEP (teacher)
Computer use and laboratory facilities	NAEP (school)
Resource adequacy	NAEP (teacher)
Computers available at the school*	
Experienced teachers' salaries*	
Teacher characteristics	
Teacher descriptors (age, race/ethnicity, gender, educational background, certification status, employment status, assignment)	NAEP (teacher)
Experience (years taught, years at this school, years at schooling levels)	NAEP (teacher)
Comfort with subject matter	NAEP (teacher)
Recency of educational enrichment**	
Student characteristics	
Descriptors (race/ethnicity, gender, parents' education, parents' occupation at the senior high level)	NAEP (student)
Courses taken and grades (high school)	NAEP (student)
SES—more complete individual measures*	
School characteristics	
Math and science course offerings—develop better lists*	
Course-taking policies—time on mathematics and science:	
a. High school graduation requirements	NAEP (school)
b. Hours required per week—middle/jr. school	NAEP (school)
Student enrollments in math, science, and computer courses**	
Dropout rates—senior high school	NAEP (school)
Dropout rates—add to middle school/ junior high level*	
Teacher planning time	NAEP (teacher)

Table 2 (continued)

Indicator*	Data Source
Classroom characteristics—curriculum	
Textbook and materials use	NAEP (teacher)
Coverage of core topics*	
Classroom characteristics—instruction	
Teaching methods	NAEP (teacher/student)
Teaching methods—shift more detailed listing of activities from student to teacher questionnaire*	
Homework	NAEP (teacher/student)
Access to laboratory equipment and computers*	
Student use of laboratory equipment and computers	NAEP (student)
Frequency and type of student assessment**	
Student achievement	
Math assessment	NAEP (student)
Science assessment	NAEP (student)
Scores of college-bound seniors	NAEP (student)
Scores of prospective math and science majors	NAEP (student)
Student participation	
Past engagement in math- and science-related activities	NAEP (student)
Extracurricular activities—include structured school, museum and university-sponsored programs*	
Concurrent math/science course-taking**	
Student attitudes	
Interest, liking, confidence, sex-stereotyping usefulness	NAEP (student)
Perceptions of usefulness	NAEP (student)
Conceptions of nature of math and science**	
Intended college major (senior high)**	
Expectation of using science/math in career	NAEP (student)

* = Expansions of existing NAEP items.
** = Additions of new NAEP items.

V. ASSESSING THE OPTIONS

Like most policy alternatives, no one of the indicator options presented in Sec. IV is clearly superior to any other, and selecting among them will depend on the NSF's preferences about how the information it collects should be used and how much it wishes to pay. We now examine some of the issues involved in making that decision.

USEFULNESS

Figure 7 rates each option on five basic functions that an indicator system might serve. All the options would provide some descriptive information about national trends in mathematics and science education. However, largely because of their sample design, the piggyback and independent options could do so in much greater detail. For example, the piggyback option would sample ethnic minorities in sufficient numbers to disaggregate by social class and gender and could thus provide more precise information about participation in mathematics and science and about how the effects of ethnicity may be confounded by social class and gender unless analyzed separately. Also, only the piggyback and independent options could provide information about the status of mathematics and science education in individual states.[1] Our interview data suggest that because of the critical role of states in the funding and governance of education, national-level data alone would be of limited use to policymakers.

The early-warning function provided by an indicator system would allow the NSF to identify emerging trends and problems and thus focus its agenda. One can think of many trends (e.g., changes in the demand for mathematics and science teachers in response to new course requirements, sharp declines in achievement among certain student groups) that if identified early enough might be dealt with more practically and constructively than they now can be. Some aspects of schooling, such as the link between student course-taking and achievement, are so well-established that descriptive trend data are sufficient to function as an early warning system. However, when less is known about relationships, an early-warning system must document national

[1]However, if Congress accepts the recommendations of a study group chaired by former Governor Lamar Alexander and H. Thomas James (1987), the NAEP will be expanded to include state-by-state data. A patchwork option based on the NAEP could then describe the status of mathematics and science education in individual states.

Uses	Status quo	Patchwork	Cyclical studies	Piggyback	Independent
Describe national status					
Describe state status					
Early warning					
Inform policy & practice					
Provide leadership					

Not capable
Does not provide information on most indicators in the comprehensive model and/or does not permit the linking of indicators

Partially capable
Provides information on some indicators in the comprehensive model in a manner that permits them to be linked

Mostly capable
Provides information on most indicators in the comprehensive model in a manner that permits them to be linked

Fig. 7—Uses of indicator options

or state-level trends and relate different domains of the system with others (e.g., by documenting the links between curriculum content and student outcomes). The status quo and patchwork options enable only limited analyses of the relationships among the various domains of schooling. Cyclical studies, on the other hand, by virtue of their in-depth focus, can address the issue of analytical links; however, their smaller sample sizes may not produce nationally or state-representative data. Consequently, only the piggyback and independent options are currently capable of completely fulfilling this function.

The indicator system could help in identifying policies that appear to succeed in improving mathematics and science education. Such findings might serve as hypotheses for exploring, evaluating, and perhaps initiating new policies and programs, or in modifying existing ones. However, to target resources effectively or to modify policies, policymakers need information about unintended consequences and potential problems at the level of school districts or even schools and classrooms. For that reason, this function requires data that allow the links among schooling domains to be analyzed and that permit distinctions to be made at least among different kinds of districts and schools (e.g., urban vs. rural, poor vs. affluent). Although the status quo and patchwork options may provide information that will help in suggesting general directions for policy, the other three options would provide more precise data that would allow conceptualization of alternative policies and practices.

If it chose to do so, the NSF could use its indicator system to provide leadership in the reform of mathematics and science education. This would imply moving beyond just reporting existing data to more comprehensive measurement of education. Because schools will be eager to "look good" on indicators, these measures might motivate them toward the types of school processes and outcomes that the NSF believes would improve the quality of science and mathematics education. For example, the NSF might seek to influence the elementary school science curriculum by reporting on the extent to which students are engaged in "hands-on" science learning, and then analyzing the links between such activities and later choice of high school courses. Only the options that allow for the design of new measures permit an indicator system to be used in this way.

In sum, only the piggyback and independent options are fully capable of meeting all the major uses an indicator system might serve. The other three options are particularly deficient in their ability to describe the state-by-state condition of mathematics and science education and to provide leadership in pushing the frontiers of indicator development and, in turn, providing direction for fundamental policy

change. However, as we will see below, the tradeoff is that the less-useful options are generally more feasible to implement.

FEASIBILITY

Difficulty or ease of implementation includes the five dimensions shown in Fig. 8, as well as issues of cost. The first dimension is the question of how much technical expertise is required to design and operate an indicator option and to interpret the data it generates. As the complexity of the option increases, from the status quo to the independent option, the difficulty of use likewise increases. With each option, the sample design, the underlying psychometric theory, and the statistical analysis become more complicated. For example, analyzing relationships among schooling domains would probably require linking data from multiple instruments. No secondary analyst has yet attempted that using the Educational Testing Service (ETS)-NAEP data. Similarly, the item sampling scheme that ETS introduced to make the NAEP more efficient also makes it more difficult to analyze. Although the five options vary significantly in the extent of analytical expertise required, even the most highly skilled analysts are currently unable to measure all important features of schooling reliably and validly. Insufficient technical expertise is thus a problem, no matter which indicator option is selected.

Respondent burden is increasingly becoming an issue as indicator systems proliferate. It is not uncommon for students in large districts to be tested by the state and also by the local district and then be required to participate in other data collection activities such as the NAEP. If the NSF were to operate its own independent system, the aggregate data burden would be further increased. This would be particularly true for teachers and students in urban areas because of the need to gather state-representative data and to oversample ethnic minorities. Even the piggyback option presents data burden problems. For example, the NAEP has placed a firm limit of one hour on both its assessment and the background questions it asks of students. If the NSF were to piggyback additional items onto the NAEP, the one-hour limit would probably have to be violated.

Political support means not only that sufficient resources are available from the NSB and Congress, but also that state and local agencies cooperate with data collection. In large part, that cooperation will depend on whether state and local officials perceive themselves as benefiting from the information reported. As the information produced becomes more comprehensive and disaggregated to the state level and

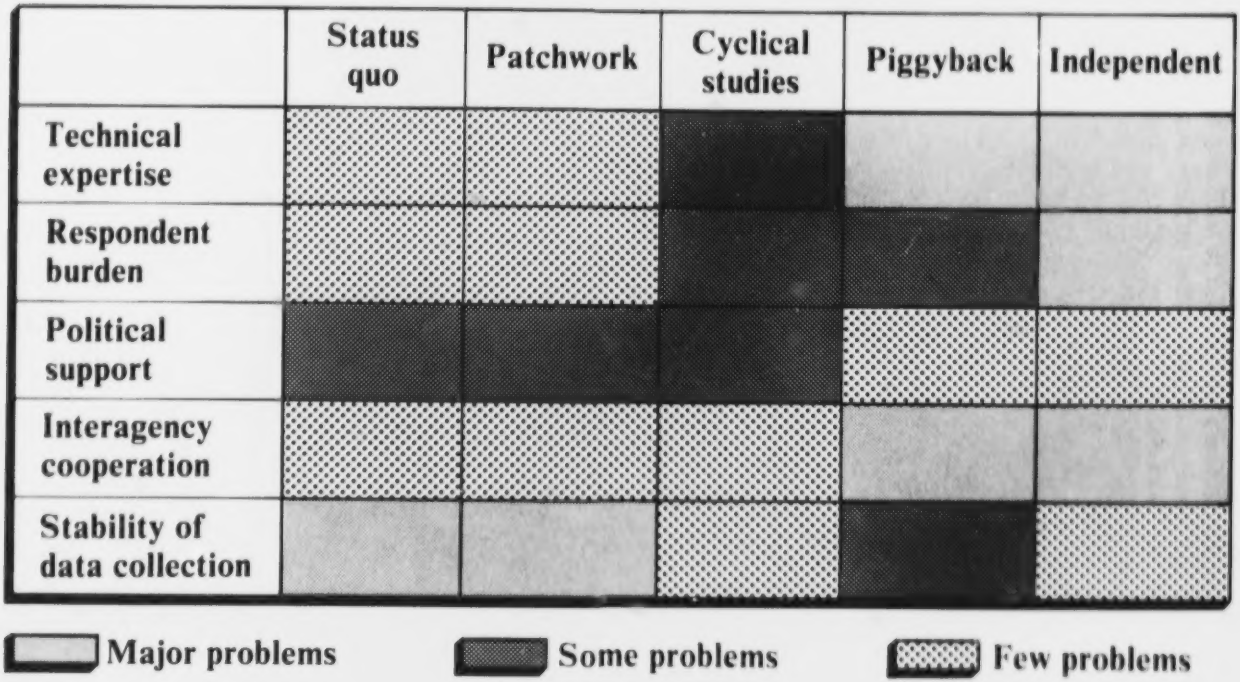


Fig. 8—Feasibility of indicator options

by district type, political support is likely to increase. A piggyback or independent system may generate information in sufficient detail to be of use to federal and state policymakers, but political support from local policymakers is questionable for these options. Support from local officials will depend on their perception of the value of information disaggregated by district type (but not by specific districts), as compared with the burden of data collection. The other three options will be of greatest use to those interested in national trends and of only limited interest to policymakers at lower governmental levels.

Interagency cooperation is not a problem if the NSF uses only data already being collected by another agency or if it collects data itself through small-scale cyclical studies. However, most piggyback options will depend on extensive cooperation with the ED and compatibility with its priorities. Any continuing relationship between the two agencies will require a division of responsibility in the development of new items, sample design, the allocation of costs for data collection, and the analysis and reporting of data. On the other hand, if the NSF were to decide on an independent strategy, it might find itself in direct conflict with the ED which has primary responsibility for the collection and reporting of educational statistics.

Somewhat juxtaposed with interagency cooperation is the stability of data collection over time. To the extent that the NSF has sole responsibility for an indicator system and does not have to rely on others for data, stability is not an issue. However, it can be a serious limitation for the status quo, patchwork, and piggyback options. For example, although the NAEP has been relatively stable over time, there is no guarantee that the ED's current priorities will continue or that they will remain consistent with those of the NSF. Even within the relative stability of the NAEP, key items and the grade levels assessed have changed, creating technical difficulties that may limit the amount of time-series data that can be generated.

A final feasibility issue is the cost of different indicator options. Table 3 summarizes cost ranges for the five options, and Appendix B details how those costs were estimated. The $40,000 per year estimate for the status quo option represents the commissioning of several special analyses of existing data as contributions to the precollege chapter of *Science Indicators* (it does not include the cost of NSF staff time to prepare that chapter). The costs for the patchwork option represent the resources needed to analyze data from about eight existing sources within the framework of an integrated model of schooling. Such an analysis would focus on many of the topics now included in *Science Indicators*, but would provide a more comprehensive and detailed analysis of major trends and the factors contributing to those trends.

Table 3

COST OF INDICATOR OPTIONS

Option	Cost per Year
Status quo	$40,000
Patchwork	$150,000
Cyclical studies	$500,000–$1 million
Piggyback	$2.5 million–$6 million
Independent	$23 million–$34 million

The costs for cyclical studies will vary, depending on the topic and the research methods used. For example, a teacher survey would cost about $500,000; a detailed curriculum study with a textbook analysis, about $1 million.

The piggyback option covers a wide range of possible costs. For example, if the NAEP collected mathematics and science data on the fifty states every four years, the additional costs would be $2.5 million per year, or $10 million over the four-year cycle. However, if data were collected every two years, the additional costs would increase to $6 million per year.

The independent option is the most expensive because it involves substantial development costs. The range of $23 million to $34 million represents differences in the frequency of data collection (two years vs. four), the size of the student and teacher samples, the extent of classroom observations, and whether or not any innovative types of student assessments are used in addition to multiple-choice items. (Tables B.1 and B.2 in Appendix B illustrate the costs involved in one type of independent indicator system.)

In sum, as the complexity of the option increases, from the status quo to a fully independent indicator system, technical expertise becomes increasingly important. An expansion of existing data collection programs to meet the NSF's specifications (a piggyback system) or the development of an independent system requires state-of-the-art breakthroughs in sampling, measurement, and data analysis. Likewise, the independent option increases respondent burden because it adds yet another data collection activity in the schools. Political support tends to increase with more comprehensive indicator systems because the enhanced monitoring capability can more closely address policymakers' concerns and suggest possible targets for reform. However, respondent burden and political support generate very real tradeoffs,

particularly at the local level: Support depends on policymakers' perceptions of the usefulness of monitoring information as compared with the overall burden the indicator system imposes. The difficulty of interagency cooperation plays an especially important role in the piggyback option; the distribution of authority across agencies also becomes an issue with the independent system. The independent option is the most likely to be stable over time, since it would be under the direct control of the NSF. Finally, costs vary substantially, not only across the five options, but also among the various alternatives included in the cyclical, piggyback, and independent categories.

VI. RECOMMENDATIONS AND NEXT STEPS

Clearly, no one of the five generic indicator strategy options can accomplish everything. Choosing one strategy or combination of strategies will involve evaluation of a variety of tradeoffs. These tradeoffs inevitably enhance or diminish the usefulness and feasibility of the systems, and also affect the impact of NSF efforts on other agencies involved in indicator development.

RECOMMENDATIONS

After assessing these tradeoffs, RAND recommends that the NSF should develop an indicator system that:

- Provides national and state-level information.
- Facilitates comparisons
  - with itself in a time series
  - with normative standards (e.g., appropriate coursework for high school graduation)
  - with data from other countries
  - across different populations.
- Builds on existing data collection efforts where possible.
- Develops new measures where adequate ones are not currently available.

Frequent Data Collection

We also recommend that key science and mathematics indicators be generated every two years. The education reform movement of the past few years has created a very fluid policy environment. As schools respond to new policies, major indicators of schooling are likely to change in significant ways. Information about the nature of those changes must be made available to the science and mathematics education community as soon as possible. Furthermore, a sound trend analysis of changes in educational practices and outcomes requires at least two, and preferably three or four, time points. A two-year cycle permits a stable time series to be established much sooner (e.g., in six as opposed to twelve years). Finally, there is always a lag between data collection and subsequent analysis and reporting. More frequent data collection will mitigate that problem.

Special Adjunct Studies

The decision to develop and field an indicator system, regardless of the particular option chosen, should carry with it a commitment to conduct additional, adjunct studies. (The relationship between these special studies and an indicator system is illustrated in Fig. 9 in the context of the NSF's mission.) These additional studies should be of two types: substantive and methodological.

Substantive studies are commissioned when the indicator system shows shifts in such aspects as patterns of student achievement and these changes appear to be associated with changes in teacher quality. Because the indicator system trades off breadth for depth in data

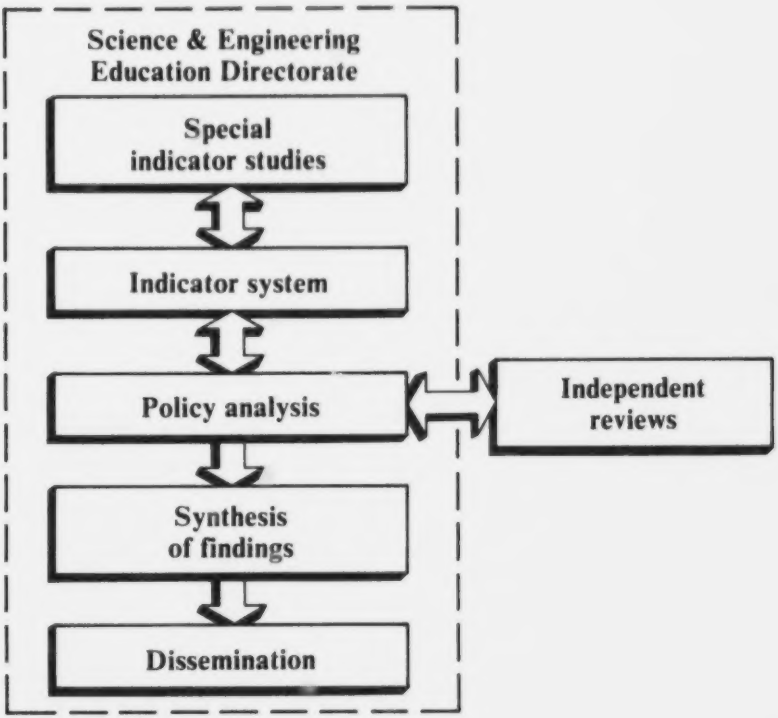


Fig. 9—Indicators in the context of NSF's broader mission

collection, the data collected will be too broad to pinpoint possible causes of observed changes or to suggest alternative policy implications. Therefore, the NSF should be prepared to commission special studies that go significantly beyond the findings of the indicator system in their depth of analysis and power of explanation.

The NSF should also be committed to conducting ongoing methodological studies of the indicator system itself. As noted earlier, the most obvious challenge to NSF may stem from the paucity of adequate educational measures. Most analysts would agree that the quality of the teaching force should be a key indicator. But while we can easily measure teachers' *qualifications* (i.e., their credentials and training, and their years of experience in the classroom), these factors do not really capture teachers' ability to teach. We lack direct indicators of teachers' skill in the classroom because of time, cost, and technical feasibility limitations. Another component with inadequate measures is student achievement, the factor many consider the most important for indicators. We have fairly good paper-and-pencil measures of the most commonly taught basic knowledge and skills, but we lack adequate measures of children's ability to think critically, to apply their knowledge, or to solve problems. A third area of needed development is curriculum. Of all the components of the education system, we probably know least about measuring those related to the curriculum students experience. Before an indicator system can be useful in assessing these and other significant components of the education system, new measures must be developed.

The Context of Analysis and Reporting

Finally, as the NSF invests in an indicator system, it should also invest in how that system can be used constructively in the political process that governs education. In addition to a commitment to special studies as essential adjuncts to an indicator system, the NSF should develop specific procedures for analyzing and reporting indicators. These processes, too, are suggested in Fig. 9. They are essential because any data emerging from the NSF indicator system will be highly visible and very important. Members of the educational community are already feeling tremendous pressure from national and state indicators. And as new systems are developed, those individuals will be judged more publicly than ever before with measures over which they have little control. As a result, educators themselves will undoubtedly exert a great deal of pressure (as they currently do in many states), attempting to affect the selection of indicators, to influence the level of

data aggregation and analysis, and to shape the methods of interpreting, releasing, and reporting data.

Yet in most cases, indicator data are unlikely to produce unequivocally good or bad news about schooling. Most of the data collected will be susceptible to various interpretations. Pressure to interpret and report data in ways that advance the interests of particular groups will be substantial. If indicator interpretations are salient, they will inevitably influence what happens in schools. In the worst cases, irresponsible indicator interpretations or reporting mechanisms might have unintended consequences and may actually retard significant improvement in mathematics and science education. Alternatively, indicators simply presented as numbers without interpretation (particularly if they are at the national level only) may be ignored.

The NSF's active involvement in analyzing the policy implications of indicator data, synthesizing findings, and developing dissemination strategies may help circumvent the irresponsible use of indicators and may promote the salience of useful and valid interpretations. For example, the NSF should not simply report time-series data on a particular indicator or set of indicators. Rather, it should analyze those time series and attempt to understand the relationship between student characteristics and the nature of the education they are receiving, and what happens over time to outcomes such as participation and achievement. Moreover, the NSF should take the technical findings of the indicator system and translate them into language understandable to a variety of audiences. By engaging in thoughtful analyses and reporting on the *full range of responsible meanings* that might be attached to indicator data, the NSF can help policymakers avoid jumping to the "one best explanation" or the "obvious solution" and can encourage deliberation about the meanings and policy implications of those data.

Finally, the analyses produced by the NSF would profit from the scrutiny of an independent review process—perhaps by an established body such as the National Academy of Sciences. Such a body could review each policy analysis coming out of the indicator system and register its evaluations or recommendations for further analytical work.

Even with these processes, NSF indicators cannot provide a single interpretation of past events, offer clear judgments about the present, or provide direct answers about what should be done next. But they will bring new knowledge to bear on educational issues, stimulate more thorough discussion and debate, and suggest creative new solutions to problems. With such an approach, the NSF indicator system can become an become important tool to aid the ongoing political dialogue about national goals for science and mathematics education and how they can best be accomplished.

Implications for the NSF's Role

The approach we have recommended would result in the NSF having much better information available for planning its own agenda, and would also give precollege mathematics and science education greater visibility nationally. This heightened attention would have important implications for the NSF as an organization. It would probably lead to a considerable expansion of the Foundation's policy role in precollege education. By disseminating more complete information about mathematics and science education and by conducting relevant policy studies, the NSF might find itself drawn into major policy debates and political controversies. The Foundation might find it more difficult to maintain its image as a neutral source of information under such circumstances, but being viewed as an active participant in major education policy debates could significantly increase the NSF's ability to shape the direction of mathematics and science education policy at all governmental levels.

Similarly, an active role in the collection and dissemination of education indicator data would bring the NSF into continued contact with a broader range of governmental agencies and constituent groups. Instead of dealing primarily with mathematics- and science-related groups, as it does now, the Foundation might find its relevant public expanded to include state and local policymakers and interest groups concerned about education generally. Again, this change would present a series of tradeoffs. Mathematics and science education would become more visible, thus increasing its likelihood of becoming a higher priority on the nation's education policy agenda. At the same time, other NSF priorities might have less attention and resources devoted to them. However, as long as the Foundation is sensitive to these tradeoffs, it should be able to accommodate a more visible and active role in precollege mathematics and science education within its existing organizational structure and priorities.

NEXT STEPS

Despite these recommendations, the uncertainties in the current environment suggest that the NSF should postpone developing a full-blown indicator system. For example, the ED is currently revising its statistical reporting system. The recent report of the task force on the NAEP has recommended expanding the current assessment to include state-level data (Alexander and James, 1987). The ED is also considering the implementation of an Integrated Elementary and Secondary Data System that would sample and report indicators on about 4,000

schools annually. In addition, the Council of Chief State School Officers is working to define indicators commonly across states and to collect data at the same time to generate reliable state-level indicators. If these efforts are successful, they will provide a major source of education data for the NSF. They may also create promising vehicles for the piggybacking option. Therefore, we recommend that the NSF spend the next two or three years developing its capacity to produce and report indicators; working with other agencies as they develop more concrete strategies; and considering further what the most productive options will be in the future.

Three activities could be undertaken immediately to move the NSF toward these goals: (1) initiation of efforts to develop better indicators of science and mathematics education using existing data, i.e., building a patchwork system; (2) initiation of studies to develop better indicators in areas where current measures are nonexistent or inadequate; (3) sponsoring of exploratory studies of how indicator data might be made more useful to policymakers.

Building Indicators from Existing Data

The NSF can immediately develop indicators for mathematics and science education by constructing a patchwork model of the type described in Sec. IV. Although a piggyback system appears more desirable, the immediate demand for an integrated system of indicators leads us to propose the building of a patchwork indicator system as a logical next step for the NSF. (This suggestion is also supported by the results of our research and consultation with our advisory board.) Since a patchwork system requires no new data collection, it is relatively inexpensive and could be implemented immediately. Perhaps more important, development of a patchwork system is not contingent on the decisions of other agencies. Consequently, this approach would permit the NSF to move forward on its indicator system without investing resources in new data collection efforts that may be duplicated in the future.

Implementing a patchwork indicator system will require an assessment of what indicators can be produced from existing data systems, what policy-related questions can be addressed with these data, and what additional data would be needed to produce a comprehensive monitoring system. We recommend that the patchwork system be guided by both a conceptual model of the education system and the utility and validity criteria identified earlier. We recommend that the patchwork system initially use the best available data from a variety of sources to produce indicators of:

- Teacher quality.
- Curriculum and instructional processes.
- Student participation in mathematics and science education, including the distribution across different types of students and levels of scientific knowledge.
- Achievement outcomes.

Where possible, the analyses should explore relationships among the major components of schooling—school organization, student characteristics, teacher qualifications, curriculum, instructional processes, and student participation and achievement outcomes. For example, using the ED school and teacher surveys, the NSF could examine the relationships among school characteristics and teacher qualifications; using the 1985–86 NAEP data, the NSF could examine the relationships among school and student characteristics and student achievement.

In addition, the NSF might be able to take advantage of recent efforts to make educational indicator data more comparable (e.g., the addition of several NAEP items to state-sponsored assessments as the result of the Council of Chief State School Officers assessment project). These databases could allow the NSF to begin to disaggregate some of the indicators for which policymakers are currently demanding state-level information.

This first step would go as far as possible with existing data to describe the current condition of precollege mathematics and science education. Although it would focus on many of the same topics now included in the precollege section of *Science Indicators* (the status quo option), it would provide a more comprehensive and detailed analysis of major trends and their distribution over time and across different types of students. It would also explore some relationships among the major components of schooling.

This exercise would lead to more concrete recommendations for the operational phase of a science and mathematics indicator system, and an assessment of the extent to which indicator system goals can be met using a patchwork of disparate data sources.

New Indicator Development

The value of any monitoring system depends on the quality of the indicators that comprise it. These indicators must be, above all, reliable and valid. Yet, as we have noted throughout this report, perhaps the single greatest barrier to the design of any educational monitoring system is the lack of such indicators. Several domains of the indicator system warrant intensive developmental work because of validity

problems. Of particular concern are current gaps in our ability to collect information about student achievement in science, teacher quality, and curriculum quality—key indicators that we had to exclude from our core set of indicators because of measurement problems. As the NSF begins to develop indicators from existing data, it should also be identifying new developmental work and data collection efforts that would improve the quality of indicators.

Indicators of Student Achievement. Achievement indicators, aggregate statistics based on achievement-test scores, are the centerpiece of any educational indicator system. They are typically based on verbally loaded, multiple-choice questions that have been designed to accommodate the range of science curricula in schools across the country. With this broad, cost-efficient approach to testing, these tests (e.g., the SAT, NAEP, or HS&B tests) are better measures of reading ability, verbal ability, and abstract reasoning than of subject-matter achievement (even when test material "looks like" science). For example, students holding misconceptions about the concepts of "force" or "gravity" may still answer multiple-choice questions correctly, either for the wrong reason or because the students are "test wise."

One alternative is a performance test that the ETS is now developing for the NSF. This test presents students with a set of laboratory apparatus and asks them to perform a scientific task. Although such an approach may measure aspects of scientific achievement not tapped previously, it is difficult to administer and score, and it is also quite costly. Experience with "hands-on" performance testing in education and the military indicates that this technology is more than five years away from becoming cost-efficient for large-scale data collection.

Another alternative is to explore a middle ground between the traditional multiple-choice test and the more expensive, less reliable performance test. This might consist of a two-stage project: The first stage would improve pencil-and-paper technology and would have an immediate impact on the quality of achievement indicators. For example, the NSF might develop achievement indicators that measure content knowledge in areas where the underlying concepts are generally acknowledged to be an appropriate part of the science curriculum (e.g., gravity, force) and where research shows the nature of students' misconception of those concepts. The second stage would explore microcomputer technology in the form of complex simulations of science experiments as a means for testing "scientific thinking."

Teacher Quality Indicators. One of the most difficult tasks in the design of educational indicators is the development of reliable and valid aggregate measures for variables that have traditionally been

measured only in small-scale experimental or observational studies, or have been incompletely specified at the aggregate level. Teacher quality is one of these variables.

Policy-related discussions about the condition of mathematics and science education almost inevitably revolve around assessments of teacher quality. Yet because of a lack of more valid indicators, the policy community must rely on limited measures such as teacher certification, because no measures are available that are more consistently related to student achievement. More detailed and generalizable measures of teacher qualifications are needed, along with indicators of the relationship between teacher quality measures and teaching quality measures.

To provide better indicators of teacher quality, developmental work is needed in three areas:

- Development of more sophisticated measures of teachers' education and work histories as an alternative to certification status as an indicator of teacher qualifications.
- Identification of teacher characteristics that influence teaching practices in science and mathematics.
- Identification of those school factors that constitute the enabling conditions of good teaching.

Composite measures that capture both personal qualifications and the conditions that facilitate good teaching will provide more powerful indicators of teacher quality than we currently have.

Curriculum Indicators. Curriculum indicators are central to any mathematics and science indicator system, but they are among the most difficult to design. There is a lack of consensus about the "ideal" curriculum, and curricula vary greatly as a result of the historical right of states and localities to determine their own curriculum. Moreover, few empirical studies have been made of curriculum effects, and it is extremely difficult to frame curriculum measures that will provide useful comparative information.

Within these constraints, however, the NSF could develop indicators that move beyond the current emphasis on content coverage. Researchers could build on the results of smaller observational and analytic studies that capture aspects of curriculum such as the coverage and accuracy of textbooks and the process by which students are instructed. The NSF might focus on ways that curriculum indicators could better measure the scientific accuracy of what is taught and the extent to which instruction reflects modes of scientific inquiry. It

could also explore the match between various curriculum approaches and the types of student being taught.

The purpose of such work would be to develop a methodology for collecting curriculum indicator data where none currently exists, and then to test that methodology in a small number of school districts.

Producing Useful Policy Information

Since the emergence of the social indicators movement in the 1960s, two new bodies of policy research have addressed the shortcomings of that effort. One focuses on the study of knowledge utilization by policymakers (e.g., Lynn, 1978; Lindblom and Cohen, 1979; Weiss, 1978), and the other concerns the development of policy indicators (e.g., de Neufville, 1985; MacRae, 1985; Murnane, 1987). The NSF could also apply the findings of that research to available data on mathematics and science education to assess whether an indicator system can actually produce useful information for policymakers.

The NSF could examine the extent to which an indicator system can address a variety of common policy issues (e.g., the relative effectiveness of modifying different schooling components, or the opportunity costs associated with concentrating resources on one component over another). Such work could focus on pressing policy concerns, such as the need to increase the number of students in the scientific manpower pipeline. It could outline the specific policy questions that need to be analyzed and then assess the extent to which various indicator system options can address those questions. These efforts could also be coordinated with efforts to develop achievement, teacher quality, and curriculum indicators to make certain that those indicators are useful to the policy community.

OPPORTUNITY, PROMISE, AND CAUTIONS

Two critical issues have emerged as high priorities for policymakers: (1) Is precollege education providing adequate levels of scientific literacy to ensure a productive workforce in an increasingly technological environment? (2) Are schools producing an adequate supply of prospective scientists and mathematicians to increase (or even maintain) the nation's competitiveness? In an era dubbed "the information age," it is not surprising that these concerns are accompanied by a renewed demand for educational indicators. Recent advances in information technology have made it possible to amass and analyze infor-

mation with an accuracy and speed that could not even be imagined several decades ago.

The current demand for better information about the quality of schooling provides a welcome opportunity for the NSF. However, currently available measures of schooling, despite improvements over the past twenty-five years, remain underdeveloped. We rely, for the most part, on a rather narrow and limited set of outcome measures (e.g., standardized multiple-choice tests of achievement, and graduation rates). Consequently, the NSF has an opportunity to develop different and better measures of important schooling processes and outcomes.

An indicator system may contribute to the understanding of the educational system in three ways: (1) by describing the system's performance; (2) by providing data that generate hypotheses for explaining trends in outcome indicators; and (3) by better informing educational policy (i.e., by enabling policymakers to pinpoint specific problem areas and to more precisely target improvement efforts). As indicator development proceeds, the collection of less-than-perfect school data is likely to spur efforts to improve our understanding of the school context and our capacity to measure its central features. A focus on the indicators themselves may contribute to the development of more sophisticated measures of the features of schooling that are most valued by policymakers, educators, and the public.

Nevertheless, even well-developed indicators, based on a comprehensive model of the education system, cannot be used to explain the causes of schooling outcomes. They cannot possibly provide the complex and interactive database necessary for fully understanding the relationships among schooling components. Neither will they point directly to solutions for educational problems. In the past, the promise of social indicators outstripped the reality. Our understanding of social systems and means of social measurement was inadequate to specify valid and reliable indicator models. Education indicator systems are similarly vulnerable.

The overarching issue facing the NSF in developing an indicator capacity is well-known to scientists: how to be both comprehensive and parsimonious. How can the complexity, subtleties, and diversity of schooling be represented in ways that permit the collection, interpretation, and communication of valid and useful information about science and mathematics education? With the limitations of educational technology clearly in mind, the NSF and other indicator developers should proceed with caution, taking care to elaborate fully on shortcomings, questions, and issues not fully explored, and always aware of what educational indicators can and cannot be expected to do.

Appendix A

CENTRAL FEATURES OF SCHOOLING

Listed below are the variables and potential indicators identified from past research as central features of schooling. They are discussed in Sec. II.

RESOURCES

Per-Pupil Expenditure

A1A Annual per-pupil operating expenditures (state and district)

Tax Effort

A2A Percent of personal income spent for education

Teacher Salaries

A3A Average teacher salary (state and district)
A3B Beginning teacher salaries (state and district)

Parent/Community Education Level

A4A Average adult educational attainment
A4B Average parent educational attainment

Teacher Resources Per Pupil

A5A Average class size (district)
A5B Average teaching load (district)

Materials, Equipment, and Facilities

A6A Annual material and equipment expenditure per pupil (state and district)
A6B Annual capital expenditure per pupil (state and district)

Specific State Resource Appropriations to Math and Science

A7A Annual specific state resources to M & S (state)

SCHOOL QUALITY

Access

- B1A Instructional time—schoolwide (elementary and secondary)
- B1B Qualified teaching staff
- B1C Staff attitudes
- B1D Resource adequacy
- B1E Facilities
- B1F Expenditures on math and science education
- B1G M & S course offerings (secondary schools)
- B1H Instructional time in M & S (elementary schools)
- B1I Use of M & S specialists (elementary schools)

Press, Emphasis

- B2A Student attendance rates
- B2B Student perception of press in school climate
- B2C Teacher or administrator perception of press
- B2D Opportunities for schoolwide recognition
- B2E Student interest in M & S careers
- B2F Student enrollment in M & S courses
- B2G HS graduation requirements in M & S

Teaching Efficacy

- B3A Salaries of M & S teachers
- B3B Teaching load and class size in M & S
- B3C Demands on time of M & S teachers
- B3D M & S teachers perceptions of working conditions
- B3E Teacher perception of problems that interfere

CURRICULUM

Policies

- C1A Goals
- C1B Content—breadth and depth of coverage
- C1C Mode of presentation
- C1D Textbook and material policies
- C1E Assessment
- C1F Sequencing & pacing of lessons & courses

Classroom Practice

- C2A Goals
- C2B Content coverage

C2C Mode of presentation
C2D Textbook and materials use
C2E Frequency and nature of student assessment
C2F Grouping practices

Quality

C3A Congruence of curriculum with expert judgments
C3B Scientific accuracy of curriculum
C3C Pedagogical appropriateness of curriculum in M & S

INSTRUCTIONAL PROCESS

Instructional Resources

D1A Textbook topic coverage, quality
D1B Pupil/teacher ratio, and pupil aides
D1C Teacher planning time/in-service opportunities
D1D Total amount of instructional time
D1E Access to laboratory equipment and computers
D1F Textbook stress on understanding, problem solving
D1G Correctness of textbooks
D1H Buffering from noninstructional tasks
D1I Teaching assignment

Instructional Policies/Processes/Activities

D2A Content coverage
D2B Amount and type of homework by topic
D2C Teaching methods
D2D Student in-class activities
D2E Flexibility of concept interpretations
D2F Opportunity for clarification
D2G Goals, methods

Instructional Climate/Management

D3A Management efficiency
D3B Amount of classroom time in active teaching
D3C Perceived difficulty, satisfaction, atmosphere

Aptitude/Prior Achievement

D4A Prior achievement

TEACHER QUALITY/QUALIFICATIONS

Enablers

- E1A Verbal ability
- E1B Experience
- E1C Recency of educational enrichment
- E1D Interest in and comfort with subject matter taught
- E1E Flexibility
- E1F Sense of efficacy
- E1G Enthusiasm

Descriptors

- E2A Teacher descriptors

OUTCOMES

Participation

- F1A Adult participation
- F1B Extracurricular participation
- F1C Voluntary course participation

Achievement

- F2A Mathematics achievement
- F2B Science achievement
- F2C Computer technology achievement

Attitudes

- F3A Self-confidence in M & S
- F3B Perceptions of M & S as disciplines
- F3C Approaches to problem solving
- F3D Perceptions of usefulness of M & S

EQUITY

Equity in Pool of Prospective Scientists and Mathematicians

- H1A Course-taking in M & S
- H1B Prospective college majors in M & S
- H1C End of HS achievement in M & S by SES
- H1D End of HS college entrance exam scores by SES

Equity at Critical Junctures of the Precollege Pipeline

- H2A Achievement in science and math
- H2B Self-selection factors among groups
- H2C Student attitudes and plans for careers in M & S
- H2D Course-taking in M & S among SES groups

Equity in Science, Math and Technology Learning Opportunities

- H3A Expenditures on math and science education (state and district)
- H3B Instructional time in science and math (elem)
- H3C Availability of labs, computers, other facilities
- H3D Course offerings in M & S (high school)
- H3E Qualified teaching staff

School Interventions to Promote More Equitable Achievement and Participation

- H4A School interventions to promote equitable achievement

Appendix B

COST ANALYSIS

INTRODUCTION

This appendix describes the procedures used to develop the cost estimates presented in Sec. V. Based on the comprehensive model of the education system described in Sec. II, 104 variables were identified that can be combined in a number of ways to provide the indicator information desired. Estimating the cost of collecting data about each of the variables first required that a data source to measure each variable be identified. The data needed for some variables are readily available from existing sources such as the NAEP, the ED's Center for Statistics, and other national surveys and research efforts. For some of the other variables, data might be collected by adding items to existing surveys of school administrators, teachers, and students. Finally, some of the variables can only be developed by collecting data through new survey instruments. Consequently, we considered three strategies for collecting the necessary data about each variable and analyzing these data as indicators:

- Patchwork
- Piggyback
- Independent

Patchwork indicator development would proceed largely through the secondary analysis of existing data collected by other national educational studies. The costs associated with the patchwork variables are largely determined by the amount of time it would take to develop an analysis plan, acquire and read data tapes, and create the variables needed for the analysis.

Piggyback indicators would be produced by adding items to generate new data from existing data collection efforts. In addition to the cost of analyzing the data collected, the piggyback option incurs the added costs associated with the development, pretesting, distribution, collection, and data entry of the items added to the surveys. Piggyback costs can also include the cost of increasing the sample size of existing surveys.

The independent option assumes the development and implementation of new national surveys and other data collection efforts to gather

information currently not collected and information not available through existing data collection efforts. If new data collection efforts were undertaken, it would be possible to eliminate or substantially reduce the scale of the piggyback option because many of the variables that could be collected through a piggyback effort could also be collected through the new data collection effort. An independent data collection effort would also result in a cost reduction for the patchwork system. Nevertheless, this is the most expensive of the three strategies.

The balance of this appendix describes the methods used to estimate the cost to collect and analyze data with the patchwork, piggyback, and independent strategies. It outlines the assumptions made and the models used to aggregate the cost estimates from individual sources into a total cost figure.

METHODS

The cost analysis was conducted in three specific stages:

1. Matching variables with data sources.
2. Determining data collection and analysis costs.
3. Aggregating individual variable costs into total strategy costs.

Matching Variables

RAND's research team identified a total of 104 variables for possible inclusion in the indicator system. These variables were divided into seven schooling domains and categorized into three priority levels.

The first task was to determine how the data to describe each variable could be collected. The variables were sorted into three groups: those that could be described with data collected through a patchwork system, those that could be described with data collected through a piggyback system. and those that required data collection through an independent system. Initially, all variables were placed in the least expensive data collection strategy. In other words, if data were available through patchwork strategies for a given variable, that variable was placed in the patchwork category. Later, as alternative data collection strategies were tested, these assumptions were relaxed. Within those strategies, the best data source was determined. Thus, the variables were sorted on four characteristics: domain, priority, collection methodology, and source.

Because data about many of the variables could be collected from multiple sources, the next step was to assign each variable to the best

source. This list was then modified to minimize the number of data sources needed. Wherever possible, we eliminated those sources that could provide information for only very few variables, if those variables could be analyzed with data from another source. Other modifications were made as additional information was collected throughout the study. For example, if data to describe a variable from one source turned out to be very difficult to analyze, and data from another source were straightforward to analyze, the source was changed to take this into account.

The cost database allowed changes to be easily incorporated in one master table as we considered different data collection strategies and different combinations of variables to include in the indicator system. PARADOX database software was used to develop a process for keeping track of all the variables.

Once a method for keeping track of the variables had been developed, the next stage of the analysis was to determine the costs of collecting and analyzing the data needed to describe each variable.

Determination of Costs

The first step of this stage of the cost analysis was to contact each of the data sources for the patchwork system. These sources included the following organizations:

- National Assessment of Educational Progress (NAEP)—Educational Testing Service (ETS)
- U.S. Department of Education, Center for Statistics
  - Longitudinal studies
    - National Longitudinal Study (NLS)
    - High School and Beyond (HS&B)
    - National Educational Longitudinal Study of 1988 (NELS '88)
  - Other studies
    - Common core of data
    - Elementary/Secondary Integrated Data System (ESIDS)
    - Survey of public and private schools
- The U.S. Census Bureau
- International Association for the Evaluation of Education (IEA)
- Education Products Information Exchange (EPIE)
- National Survey of Science and Mathematics Education (NSSME)
- Council of Chief State School Officers (CCSSO)

- National Education Association (NEA)
- College Entrance Examination Board (SAT)
- American College Test (ACT)

Each contact person was asked about the availability and format of data that might be used to develop specific variables. After the availability of the data was determined, we then estimated the costs of obtaining, analyzing, and reporting the data as part of the indicator system.

We consulted with members of RAND's Computer Information Services Department to estimate the effort required to conduct the analyses. We then used RAND budgeting techniques with assumed salary rates for various categories of personnel to calculate the costs of various analyses. For each source, a model was developed showing the personnel time and costs, as well as the estimated computing expenditures for analyzing the data. Reporting costs were added to these estimates to develop the costs of a patchwork system.

In estimating the costs of a piggyback system, we assumed that it would be possible to gain cooperation from other agencies to add items to their existing survey instruments. The costs of adding items were estimated on the basis of information provided by the relevant organizations and by RAND's Survey Research Group. This information was aggregated into a model that estimated the costs of a piggyback data collection system and provided an estimate of the savings to the patchwork system that would result from shifting some of the variables from the patchwork to the piggyback system.

Finally, the costs of each component of an independent system were estimated using published data on large surveys, information from RAND's Survey Research Group, and estimates of costs provided by the sources listed above. Again, the savings to the patchwork and piggyback options were determined, and the total estimated costs of the system were aggregated through a cost model (described below). All cost estimates were expressed in 1986 dollars.

ASSUMPTIONS

We made a number of assumptions specific to individual data sources or variables such as survey sample sizes, the level of detail required in an analysis, and the level of effort required for development of new instruments. These assumptions are described in a detailed cost memorandum that was submitted to the project staff. However, a number of general assumptions were also required for the analysis. They are outlined below.

Inflation

All dollar figures were presented in current (1986) terms, and all projections were based on 1986 dollars. The use of 1986 dollars enabled us to make more straightforward comparisons among alternatives. Ignoring the effects of inflation had the further advantage of allowing individual readers to apply their own projections of the rate of inflation to the figures without having to first convert back to 1986 dollars. This is not to say that inflation has no impact on the total costs of the proposed indicator system; rather, we chose to show what the costs would be in constant terms, leaving the question of determining an appropriate inflation rate to others.

Cooperation

Throughout the analysis, references are made to patchwork analysis of data collected by other organizations, and to piggybacking onto other data collection instruments. It is unlikely that access to data from other sources for patchwork analysis will be difficult to obtain. However, there may be substantial limitations to piggybacking on some data sources. For example, the NAEP has very strict time limits on the administration of its survey instruments, and it may not be possible to add all of the desired items to those instruments, even if the NAEP's governing structure were favorably disposed to do so. Another consideration is respondent burden. School administrators and teachers may be reluctant to fill out yet another national survey. Although these are important considerations in the development of the NSF indicator system, the cost analysis presented here assumes those issues can be resolved. The analysis also does not take into account the costs of respondent time to answer the questions in the survey instrument.

The costing techniques presented were designed to be flexible enough to allow for changes resulting from these difficulties. In addition, potential problems, such as technical difficulties in analyzing data tapes provided by a data source or previous delays in the availability of data from a source, are identified to alert readers to their possible impact on development of the indicator system.

THE COST ANALYSIS MODELS

The cost model developed to provide estimates of the total costs of various indicator system options uses Lotus 1-2-3 and was designed to allow for straightforward analysis of alternative assumptions and combinations of variables.

Analysis of the variables by domain was desired, yet data in a given domain were to be collected from a variety of sources. When analyzing data from any given source, an initial investment of time and effort is needed to learn how to read and analyze the data. Once this "investment" has been made, it is less difficult to do subsequent analyses of the data. Consequently, in the cost estimates that follow, two types of costs were estimated:

- **Start-up costs.** It was assumed that 25 percent of the costs of analyzing data from a given source are one-time-only, or start-up costs. In other words, if the data from the NSSME were to be analyzed for one domain, there would be certain start-up costs. However, if the data were used for two domains, the start-up costs would not recur. The model was designed so that start-up costs would be counted only once if data from a source were to be used in multiple domains. The model also allows the portion of total costs allocated to start-up to vary by data source.
- **Variable costs.** For each variable of interest, there are also costs associated with developing derived variables and conducting the analysis. In the cost models, it was assumed that variable costs represent 75 percent of the costs of analyzing the data from one source. These costs were divided evenly over each variable, regardless of domain to be analyzed from a given source. The percentage of variable costs can be varied by changing the portion devoted to start-up costs.

The first part of the model allowed the researchers to change any of the assumptions that were used to calculate the costs of the patchwork and piggyback options. For example, researchers could choose any combination of the eight domains desired, vary the start-up cost percentage by data source, vary the number of days required for a given analysis task as well as the estimated computer charges, and change the type of textbook analysis and number of text book series to be analyzed. In addition, researchers could change the assignment of variables from one source to any other source. Once all of the desired changes are made, the model recalculates the costs and prints out a summary of the estimated costs.

The second part of the model estimates the cost of collecting data, using all three approaches as appropriate to each variable. It also estimates the cost of the independent data collection effort and adjusts the patchwork and piggyback costs accordingly. Finally, the model allows researchers to vary the number of classroom observations to be

conducted and the number of state and district curricula to be reviewed.

The third part of the model estimates costs of collecting data for all of the variables through an independent data collection system. It allows researchers to revise the number of districts to be surveyed, the size of the student data questionnaire, the number of classrooms to observe, the type of textbook content analysis to perform, and the number of district curricula to review. A sample of the output from this model is displayed in Table B.1. The sample output from the assumptions section of this part of the model is shown in Table B.2.

Table B.1

ESTIMATED COST OF AN INDEPENDENT SYSTEM

Data Collection Method	Cost ($)	
	Low	High
State and district surveys		
Development	675,000	750,000
Survey costs		
States	112,500	125,000
Districts	475,000	600,000
Follow-up costs	138,600	173,250
Total	1,401,100	1,648,250
Administrator and teacher surveys		
Development	500,000	800,000
Administration	2,500,000	4,500,000
Total	3,000,000	5,300,000
Student assessment	9,800,000	16,100,000
Student characteristics		
Development	225,000	250,000
Administration	5,400,000	6,480,000
Total	15,425,000	22,830,000
Analysis of Census		
Bureau data	36,500	68,000
Classroom observation	575,000	700,000
Expert panels		
Development costs	300,000	350,000
Curriculum analysis		
State level	400,000	500,000
District level	800,000	1,000,000
Textbook analysis	450,000	600,000
Total	1,950,000	2,450,000
Analysis and reporting	833,684	1,078,517
Total	23,221,284	34,074,767

Table B.2

ASSUMPTIONS USED TO DERIVE TABLE B.1

Number of districts to be surveyed (# in C24)[a]	1000
Estimated number of data bits in student characteristics questionnaire (# in C28)[a]	200
Number of classroom observations (# in C31)[a]	1000
Textbook content analysis decisions	
Enter the type of analysis desired in D38[a]	
Content = 1	
Purpose = 2	
Process = 3	1
Enter the number of series to be analyzed in D46[a]	100
Number of district's curricula to be reviewed (D49)[a] (there will be two curriculum analyses for each district)	100

[a]The Lotus 1-2-3 cell number where responses were to be entered.

Appendix C

PROJECT NATIONAL ADVISORY BOARD MEMBERS

Richard Berry
National Science Foundation

Leigh Burstein
UCLA Graduate School of Education

Audrey Champagne
American Association for the Advancement of Science

Lee Cronbach
Stanford University

Pascal D. Forgione, Jr.
Connecticut State Department of Education

Bernard Gifford
University of California at Berkeley

Norman Hackerman
Rice University

Harry Handler
Los Angeles Unified School District

Ina Mullis
Educational Testing Service

Ingram Olkin
Stanford University

Jerome Pine
California Institute of Technology

Tom Romberg
University of Wisconsin-Madison

Senta Raizen
National Academy of Sciences

Ramsay Selden
Center for Assessment and Evaluation
Council of Chief State School Officers

Appendix D

INTERVIEW TOPICS

This appendix reproduces the guide used in interviewing policymakers concerned with precollege mathematics and science education.

Brief description of the monitoring system project

(*Define* what we mean by a *monitoring system*—the regular, systematic collection of indicator data, based on a model that includes those factors related to math and science performance. This system would be designed to assess the status of math and science education nationally.

Stress that we are considering a broad range of math and science indicator systems, and want to make certain that the options we design for such a system will meet the needs of a variety of users in the policy community.)

First, I would like to get your views about the kinds of information that you think a national science and math monitoring system should provide.

1. a. What precollegiate mathematics or science policy issues have you been involved with recently in your position as (POLICY ROLE)?

   b. In considering these issues, what kinds of data or other information did you use?

      e.g., research findings and implications; aggregate data on the status of science and math education (e.g., on course enrollments, student achievement); practitioner experience or viewpoints; public opinion data; advocacy positions.

   c. Are there any other types of information that you would have liked to have, but did not have available?

2. In your position as (POLICY ROLE), what *types of information* would you most like to receive from a national monitoring system?

   PROBES: any other information needed, beyond just achievement indicators (e.g., on resources, teacher and teaching quality, curriculum, equity of participation)

3. How would you typically use monitoring system information?

   e.g., —to set broad policy directions for your agency
   —to help in making funding decisions
   —to provide information to your constituency or
   —the general public
   —to stimulate improvement of science and math education

4. In talking with potential users about a monitoring system, most mention three basic *purposes* that such a system might serve. These are:

   - describing the current status of precollege mathematics and science education and documenting general trends in the field
   - serving as an accountability mechanism by reporting data about how well schools and students are performing
   - improving policy and practice by providing information about which approaches seem to be working

   a. How would you rate the importance of each of these purposes?

   b. Are these appropriate purposes for NSF to address?

   c. Are there any other purposes that a monitoring system might serve that you think are important and appropriate for NSF?

5. In designing a monitoring system, do you think that NSF should establish *goals* against which math and science performance can be measured?

   IF YES: —how should that process be structured?
   —who should participate in defining those objectives (e.g., anyone outside NSF; outside the federal government)?

6. a. Who do you see as the major audiences for monitoring system results?

   PROBES:
   - those within NSF
   - people outside NSF at the federal level (i.e., Congress, Executive Branch, interest groups)
   - state and local policymakers
   - the general public

   b. Among these various audiences, which are the *most critical* for NSF's mission?

   c. How can monitoring system results be most effectively communicated to these different audiences?

   PROBES:
   - format (e.g., detailed written report, written summary, oral briefing, etc.)

7. a. One of the issues that continues to arise when we discuss the design of a national monitoring system is the *level at which data should be collected* and analyzed? Some people believe that knowing broad national trends is sufficient, while others believe that information needs to be disaggregated, at least to the level of individual states. The latter argument is usually made because so much math and science education depends on state and local policy. What do you think is the appropriate level for NSF monitoring system data to be collected and reported?

   b. Do you foresee any problems with the federal government collecting such data?

   e.g., —concerns that such data collection may lead to a national curriculum
   —overlap with existing federal, state, and local data collection efforts

8. a. As you probably know, there are several other national data collection efforts that provide some information about math and science education (e.g., NAEP). To what extent do you think NSF should try and rely on these existing efforts or should it design a system of its own that more closely addresses its particular concerns about math and science education?

b. What do you see as NSF's role in designing and maintaining a national monitoring system compared with the U.S. Department of Education?

9. a. In comparison with NSF's other activities in precollegiate math and science education, how important do you think the operation of a national monitoring system is?

e.g., —as compared with funding teacher staff development models or instructional materials development

b. Preliminary estimates suggest that a national monitoring system would cost from $3 million to $20 million a year. About how much do you think is appropriate to spend on such a system?

PROBES: • what proportion of SEE's budget should be spent on such a system?

REFERENCES

Alexander, K. L., and McDill, Edward L. (1976). Selection and allocation within schools: Some causes and consequences of curriculum placement. *American Sociological Review, 41*, 963–980.

Alexander, K. L., Natriello, G., and Pallas, A. M. (1985). For whom the bell tolls: The impact of dropping out cognitive performance. *American Sociological Review, 50*, 409–420.

Alexander, L., and James, H. T. (1987). *The nation's report card.* Report of the study group, Washington, DC: U.S. Department of Education.

Association for School, College, and University Staffing (1984). *Teacher supply/demand.* Madison, WI.

Averch, H. A., et al. (1972). *How effective is schooling? A critical review and synthesis of research findings.* Santa Monica, CA: The RAND Corporation.

Barr, R. C. (1980). "School, class, group, and pace effects on learning." Paper presented at the annual meeting of the Educational Research Association, Boston.

Barr, Rebecca, and Dreeben, Robert (1983). *How schools work.* Chicago: University of Chicago Press.

Barth, R. (1979). Home-based reinforcement of school behavior: A review and analysis. *Review of Educational Research, 49*, 436–458.

Berliner, D. (1979). Tempus educare. In P. Peterson and H. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications* (120–135). Berkeley, CA: McCutchan.

Berryman, S. E. (1983). *Who will do science? Minority and female attainment of science and mathematics degrees: Trends and causes.* New York: The Rockefeller Foundation.

Bloch, E. (1986). Basic research and the economic health: The coming challenge. *Science, 232* (4750), 595–599.

Borg, Walter R. (1980). Time and school learning. In Carolyn Denham and Ann Lieberman (Eds.), *Time to learn.* Washington, DC: National Institute of Education.

Bridge, R. G., Judd, C. M., and Moock, R. R. (1979). *The determinants of educational outcomes.* Cambridge, MA: Ballinger.

Brookover, W., Beady, C., Flood, P., Schweitzer, J., and Wisenbaker, J. (1979). *School systems and student achievement: Schools can make a difference.* New York: Praeger.

Burstein, L., et al. (1985). *Using state test data for national indicators of education quality: A feasibility study.* Los Angeles, CA: Center for Study of Evaluation, University of California, Los Angeles.

Catterall, J. S. (1986). *On the social costs of dropping out of school.* Program report series. Stanford, CA: Institute for Research on Educational Finance and Governance.

Centra, J. A., and Potter, D. A. (1980). School and teacher effects: An interrelational model. *Review of Educational Research, 50,* 273–291.

Chelimsky, E. (1986). Statement before the Subcommittee on Select Education, U.S. House of Representatives on the condition of information education. Washington, DC.

Clark, D., Lotto, L., and Astuto, T. A. (1984). Effective schools and school improvement: A comparative analysis of two lines of inquiry. *Educational Administration Quarterly, 20* (3), 41–68.

Coleman, James S., et al. (1966). *Equality of educational opportunity.* U.S. Department of Health, Education, and Welfare, Office of Education, OE-38001, Washington, DC: U.S. Government Printing Office.

Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., and Conney, T. J. (1984). *Second international mathematics study: Summer report for the United States.* IEA, U.S. National Coordinating Center, Champaign-Urbana, IL: University of Illinois.

Darling-Hammond, L. (1984). *Beyond the Commission reports: The coming crisis in teaching* (R-3177-RC). Santa Monica, CA: The RAND Corporation.

de Neufville, J. I. (1985). *Social indicators and public policy.* New York: American Elsevier.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L. S. Shulman (Ed.), *Review of research in education,* Vol. 5. Itasca, IL: Peacock Publishers.

Educational Testing Service (1987). *Learning by doing* (Report 17-HOJ-80). Princeton, NJ.

Epstein, J. L., and Becker, H. J. (1982). Teachers practices of parent involvement: problems and possibilities, *Teachers College Record, 83,* 103–113.

Fisher, C., Berliner, D., Filby, N., Marlique, R., Cahan, L., and Dishaw, W. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham and A. Lieberman (Eds.), *Time to learn.* Washington, DC: National Institute of Education.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39* (3), 193–202.

Gauld, C. (1982). The scientific attitude and science education: A critical reappraisal. *Science Education, 66* (1), 109–121.

Glasman, Naftaly S., and Bimaninov, Israel (1981). Input-output analyses of schools. *Review of Education Research, 51* (4), 509–539.

Goodlad, John I. (1984). *A place called school.* New York: McGraw-Hill.

Hall, G., et al. (1985). *Alternatives for a national data system on elementary and secondary education.* Report prepared for the Office of Educational Research and Improvement. Washington, DC: U.S. Department of Education.

Hanushek, Eric (1981). Throwing money at schools. *Journal of Policy Analysis and Management, 1* (1), 19–41.

Hilton, T. L. (1987). General problems of merging results from two or more data bases. Paper presented at the 1987 annual meeting of the American Educational Research Association, Washington, DC.

Horn, E. A., and Walberg, J. H. (1984). Achievement and interest as functions of quantity and level of instruction. *Journal of Educational Research, 77,* 227–232.

Howe, T. G., and Gerlovich, J. A. (1982). *National study of the estimated supply and demand of secondary science and mathematics teachers.* Ames, IA: Iowa State University.

Husen, T. (Ed.) (1967). *International study of achievement in mathematics: A comparison of twelve countries,* Vols. 1, 2. New York: John Wiley and Sons.

Hueftle, S. S., Rakow, S. J., and Welch, W. W. (1983). *Images of science.* Minneapolis, MN: Minnesota Research and Evaluation Center, University of Minnesota.

Johnston, K. L., and Aldridge, B. G. (1984). The crisis in science education: What is it? How can we respond? *Journal of College Science Teaching,* 20–28.

Klitgaard, R. E. (1974). *Achievement scores and educational objectives* (R-1217-NIE). Santa Monica, CA: The RAND Corporation.

Levin, H. M. (1972). *The cost to the nation of inadequate education.* Report to the Select Committee on Equal Educational Opportunity of the United States Senate, Washington, DC: U.S. Government Printing Office.

Levine, D. B. (Ed.) (1986). *Creating a center for education statistics: A time for action.* National Academy Press.

Lindblom, C. E., and Cohen, D. K. (1979). *Usable knowledge.* New Haven, CT: Yale University Press.

Locke, E. A. (1976). The nature and causes of job satisfaction. In M. Dunnett (Ed.), *Handbook of industrial and organizational psychology.* New York: Rand-McNally.

Lortie, D. C. (1975). *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

Lynn, L. E., Jr. (1978). *Knowledge and policy: The uncertain connection?* Washington, DC: National Academy of Sciences.

MacRae, D., Jr. (1985). *Policy indicators: Links between social science and public debate.* Chapel Hill, NC: The University of North Carolina Press.

Madaus, G. F., Kellaghan, T., Rakow, E. A., and King, D. J. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review, 49* (2), 207–230.

Matlack, C. (1987). Signing up science. *National Journal,* 1348–1353.

McDill, E. L., and Rigsby, L. C. (1973). *Structure and process in secondary schools: The impact of educational climates.* Baltimore and London: The Johns Hopkins University Press.

McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., and Cooney, T. J. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes Publishing Company.

Mera, H., Fallis, M., and Hallett, A. C. (1983). *Class coverage in the Chicago public schools: A study of teacher absences and substitute coverage.* Chicago: Chicago Panel on Public School Finances.

Munby, H. (1983). *An investigation into the measurement of attitudes in science education* (ED 237 347). Columbus, OH: SMEAC Information Reference Center, The Ohio State University.

Munday, L. A., and Davis, J. C. (1974). *Varieties of accomplishment after college: Perspectives on the meaning of academic talent* (ACT Report No. 62). Iowa City, IA: American College Testing Service.

Murnane, R. J. (1987). Improving education indicators and economic indicators: The same problems? *Educational Evaluation and Policy Analysis.*

National Commission on Excellence in Education (1983). *A nation at risk.* Washington, DC: U.S. Government Printing Office.

National Science Board (1983). *Educating Americans for the 21st century.* Washington, DC: National Science Foundation.

—— (1985). *Science indicators: The 1985 report.* Washington, DC: National Science Foundation.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Pascal, A. (1987). *The qualifications of teachers in American high schools* (N-2547-CSTP). Santa Monica, CA: The RAND Corporation.

Peng, Samuel S., Owings, Jeffrey A., and Fetters, William B. (1982). Effective high schools: What are their attributess? Paper presented at the Annual Meeting of the American Statistical Association. Cincinnati, OH.

Plisko, V. W., Ginsburg, A., and Chaikind, Stephen (1986). Assessing national data on education. *Educational Evaluation and Policy Analysis, 8* (1), 1–16.

Raizen, S. C., and Jones, L. V. (Eds.) (1985). *Indicators of precollege education in science and mathematics: A preliminary review.* Washington, DC: National Academy Press.

Rock, D. A., et al. (1985). *Study of excellence in high school education: Longitudinal study* (19180–82, Final Report). Princeton, NJ: Educational Testing Service.

Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking.* New York: Wiley.

Rosenholtz, S. J. (1985). Effective schools: Interpreting the evidence. *American Journal of Education, 93,* 352–388.

Rosenholtz, S. J., Bassler, D., and Hoover-Dempsey, K. (1985). *Organizational inducements for teaching.* Interim report submitted to the National Institute of Education, University of Illinois.

Rumberger, R. (1985). The shortage of mathematics and science teachers: A review of the evidence. *Educational Evaluation and Policy Analysis, 7* (4), 355–369.

Schwille, J., et al. (1981). *Teachers as policy brokers in the content of elementary school mathematics.* East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Shavelson, R. (1985). *The Dodds class of 1982: Characteristics of students remaining in the same high school, 1980–1982* (N-2319-FMP). Santa Monica, CA: The RAND Corporation.

—— (1983). Review of research on teachers' pedagogical judgments, plans, and decisions. *Elementary School Journal, 183,* 392–414.

Steinberg, L., Blinde, P. L., and Chan, K. S. (1982). Dropping out among language minority youth: A review of the literature. *Journal of Review of Education Research,* 113–132.

Walberg, H. J. (1984). Families as partners in educational productivity, *Phi Delta Kappan, 65,* 397–400.

Webb, N. M. (1980). A process outcome analysis of learning in group and individual settings. *Educational Psychologist, 15,* 69–83.

Weiss, I. R. (1978). *Report of the 1977 national survey of science, mathematics and social studies education.* Washington, DC: U.S. Government Printing Office.

Welch, W. (1984). Research in science education: Review and recommendations. In J. L. Taylor (Ed.), *Teacher shortage in science and mathematics: Myths, realities and research.* Washington, DC: National Institute of Education.

—— (1983). Research in science education: Review and recommendations. Paper presented at the Conference on Teacher Shortage in Science and Mathematics: Myths, Realities, and Research. Washington, DC: National Institute of Education.

Wharton, Y. L. (1977). *List of hypotheses advanced to explain the SAT.* New York: College Entrance Examination Board.

Wiley, David E., and Harnischfeger, A. (1974). Exploration of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher, 3,* 7–12.

Wolf, R. M. (1977). *Achievement in America.* Columbia, NY: Teachers College Press.

96 "END OF DOCUMENT"